

Short-term Forecast Model of Vehicles Volume Based on ARIMA Seasonal Model and Holt-Winters

Zhi-Hui WANG^{1,a}, Chen-Yang LU^{2,b}, Bin PU^{2,c}, Gui-Wen LI^{1,d}, and Zhen-Jiang GUO^{1,e}

¹Intelligent Center, Yunnan Science Research Institute of Communication & Transportation, Kunming, China

²National Pilot School of Software, Yunnan University, Kunming, China

^a937140839@qq.com, ^bluchenyanglcy@163.com, ^c937140839@qq.com, ^d344294117@qq.com, ^e3811247@qq.com

Abstract: In order to alleviate the urban traffic congestion and ensure traffic safety, we need to do a good job in urban road traffic safety planning, make the real-time analysis and forecast of urban traffic flow to detect changes of current traffic flow in time, make scientific planning of roads and improve the road service ability and the transport efficiency of freight vehicles. The data of short-term vehicles volume is characterized by uncertainty and timing correlation series. Given this, the ARIMA seasonal model and the Holt-Winters model are used to establish a forecasting model for the short-term vehicles volume of the city. Finally, we compare the model with predictions.

1 Introduction

City vehicles volume is one of the important basis for the preparation of urban development planning, and it is also an important basic data for economic evaluation and post evaluation of urban construction projects. For the accurate forecast of vehicles volume, it can help transport the traffic control department of scientific management, to ensure the traffic safety. Short-term forecast of vehicles volume is mainly based on a period or even the daily traffic flow changes at the start, to study the short-term variations in the flow of goods vehicles. The short-term vehicles volume is a non-stationary time series[2] with both seasonal and periodic trends[1] and with some diversification.

Because of these attributes, the two typical time series prediction models of ARIMA and Holt-Winters are used to forecast the vehicles GPS data between October 20, 2015 to November 06, 2015 of Kunming, Yunnan Province to analysis and processing.

Traditional predictive methods must collect significant amounts of data, synthesize the various factors that affect the volume of the vehicle collecting data are difficult[3]. This type of forecasting method's hard to model. Sometimes the model can be successfully established, but the model parameters are lacking the necessary data and can't be reasonably estimated[4]. Because there are two significant characteristics of urban vehicles volume: Firstly, the data varies with the seasons and has a certain periodicity; Secondly, the series formed by the volume of the urban vehicle can be treated as a random series. We can use the ARIMA and Holt-Winters prediction model to approximate the series. When the random series matches the prediction model, the future value can be predicted by

the time series and the current value[5]. Therefore, we build the forecasting model of urban vehicles volume to improve the accuracy of forecasting. Aim to plan the road traffic system scientific and improve the transportation efficiency of vehicles.

2 Data Source

The establishment of this model mainly selects the GPS data of vehicles in Yunnan Province from October 20, 2015 to November 16, 2015. Obtained the hourly vehicles volume data of 19 major cities in Yunnan Province after data preprocessing and preliminary statistical analysis. Choosing the 18 days vehicles volume data of city ID as 01 of Kunming to construct the model. The data example shown in Table 1:

Table 1. Vehicles Volume Data Example

| Data | Time Period | City ID | Vehicles Volume |
|----------|-------------|---------|-----------------|
| 20151120 | 0 | 01 | 328 |
| 20151120 | 1 | 01 | 130 |
| 20151120 | 2 | 01 | 34 |
| 20151120 | 3 | 01 | 42 |
| 20151120 | 4 | 01 | 82 |
| 20151120 | 5 | 01 | 149 |

3 Model Overview

During the period from October 20, 2015 to November 06, 2015 the volume of the total vehicle of Kunming shows a significant fluctuation. And seasonal characteristic was more significant. The seasonal cycle is the 24 hours a day.

The peak and troughs of the volume of the vehicle at each time of the day are very stable. The peak of the volume of the vehicle appears at 9 o'clock to 10 o'clock every day, the trough of the volume of the vehicle appears at 3 o'clock to 4 o'clock every day. The random error is tested by Ljung-Box test, and the P value of random error white noise test is 0, indicating that the original data has sufficient seasonal and the random error has little effect.

3.1 ARIMA Seasonal Model

ARIMA seasonal model is a form of ARIMA model. For non-stationary time series with seasonal and trendy urban vehicles volume, the stationary time series can be obtained by periodic series difference and seasonal series difference. ARIMA seasonal model can extract the time series characteristics of seasonal, trend and random errors in the vehicles volume series.

Through the timing diagram and the result of difference operation[6], the vehicles volume predict model was determined as the $ARIMA(p, d, q) \times (P, D, Q)[S]$ initially, the p is the degree of nonseasonal autoregressive[7], d is the number of differentials, q is the moving average order[8]; P is the seasonal automatic regression order[9], D is the seasonal time differences, Q is the seasonal moving average order; S is the seasonal period[10]. The complete structure of the ARIMA seasonal model is as follows:

$$\nabla^d \nabla_S^D Y_t = \frac{\theta(x)\theta_s(x)}{\phi(x)\phi_s(x)} \varepsilon_t \quad (1)$$

Among them,

$$\begin{cases} \theta(x) = 1 - \theta_1 x - \dots - \theta_q x^q \\ \phi(x) = 1 - \phi_1 x - \dots - \phi_p x^p \\ \theta_s(x) = 1 - \theta_1 x^s - \dots - \theta_1 x^{QS} \\ \phi_s(x) = 1 - \phi_1 x^s - \dots - \phi_p x^{PS} \end{cases} \quad (2)$$

Y_t is the non-stationary time series at time t , ε_t is the white noise process, S is the cycle step, the study of the vehicles volume sequence has obvious cyclical and using the hourly data as the unit of seasonal data, so the $s = 24$, x is the hysteresis operator. $\theta(x)$ is the moving average coefficient polynomial of degree q , $\phi(x)$ is the autoregressive coefficients polynomial of degree p . $\theta_s(x)$ is the moving average coefficient polynomial of degree Q , $\phi_s(x)$ is the autoregressive coefficients polynomial of degree P [11].

ARIMA seasonal model establishment steps:

- Test series smoothness. By observing the time series scatter plot, check whether the unit root exists and the ACF, PACF graph to determine whether the sequence is smooth. For the non-stationary sequence, it can be smooth by difference operation.

- Model identification and parameter fitting. According to the time series ACF graph and PACF graph to select the appropriate parameters of the seasonal ARIMA model to fit.

- Parameter estimation. Using the AIC criterion to determine the model parameters, compare the AIC values of each parameter model, select the model with the smallest AIC value as the optimal model.

- Model test. The residuals white noise test of the constructed model to determine whether the model is feasible.

- Model prediction. Use the established model to predict future time series.

3.2 Holt-Winters Model

Holt-Winters model is based on the statistical time series forecasting model, which can predict the time series with no distinct function rules but does correlate and seasonal trend. The time series of seasonal fluctuation, random variation and linear trend are decomposition, combined with exponential smoothing method to estimate the amount of decomposition and establish the prediction model.

The Holt-Winters seasonal model based on the addition is added to the horizontal component and the seasonal component[12]. The complete definition of the model is as follows:

$$\begin{cases} l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \\ s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \\ \hat{y}_{t+h} = l_t + hb_t + s_{t-m+h} \end{cases} \quad (3)$$

The t is the time of the time series. l_t is the horizontal component, b_t is the trend component, s_t is the seasonal component, m is the set seasonal period, h is the prediction period, the constant α, β, γ is the smoothing coefficient.

The Holt-Winters seasonal model based on multiplication is the horizontal component multiplied by the seasonal component. The complete definition of the model is as follows:

$$\begin{cases} l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \\ s_t = \gamma \frac{y_t}{l_{t-1} - b_{t-1}} + (1 - \gamma)s_{t-m} \\ \hat{y}_{t+h} = (l_t + hb_t)s_{t-m+h} \end{cases} \quad (4)$$

The t is the time of the time series. l_t is the horizontal component, b_t is the trend component, s_t is the seasonal component, m is the series seasonal period, h is the prediction period, the constan α, β, γ Holt-Winters seasonal is the smoothing coefficient.

4 Experimental Results

4.1 ARIMA Experimental Results

Perform the white noise test on the original sequence; The results shows the , indicate that the original sequence is a non-white noise sequence[13], which has the significance of the research.

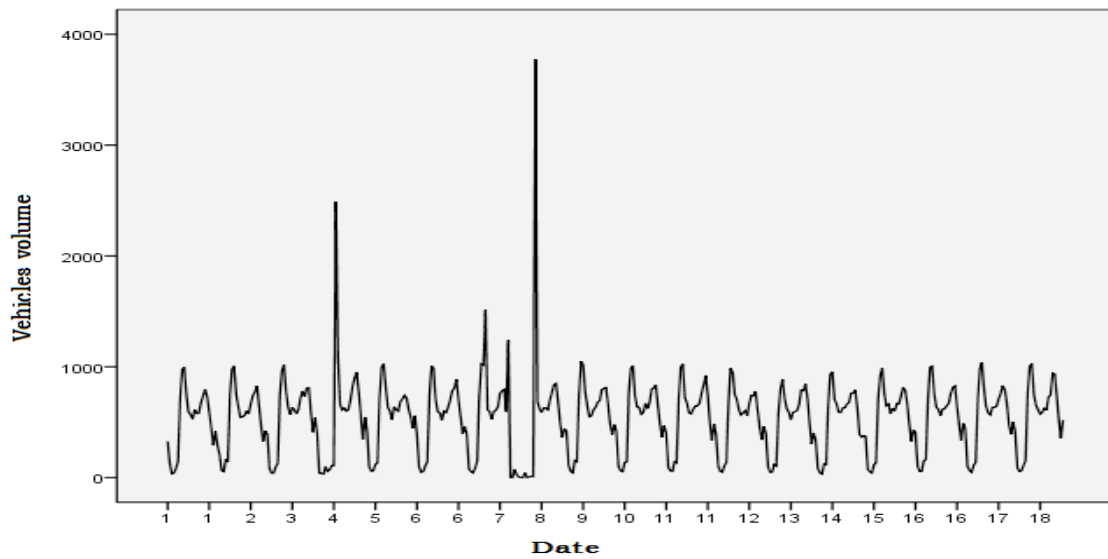


Figure 1. Original data time series graph

The vehicles volume shows the significant seasonal characteristics by per hour in Kunming. Tested by ADF, getting the probability of corresponding statistics is 0.01, Rejecting the assumption that the original sequence is a non-stationary series, the result shows that the vehicles volume time series are stationary, which consistent with the observations observed in Figure 1.

Table 2. Adf Checklist

| Dicke y-Fuller | Lag order | P-value |
|----------------|-----------|---------|
| -8.7814 | 7 | 0.01 |

By observing the autocorrelation, partial autocorrelation and ADF test result of Figure 2, the

original sequence was diagnosed as a trend stationary, seasonal non-stationary series[14,15]. So we make the one-degree difference to the series. So the $d = 0$, $D = 1$. Through the autocorrelation graph and partial autocorrelation graph, we can obtain the p and q value in the $ARIMA(p, d, q) \times (P, D, Q)[S]$. The parameter q of the $MA(q)$ can be determined by the truncation of the AC, the parameter p of the $AR(p)$ can be determined by the truncation of the PAC. From the Figure 2 we can do analysis that it could be possible the value of p could be 0 or 1, the value of q could be 0 or 1, the value of P could be 0 or 1, the value of Q could be 0 or 1.

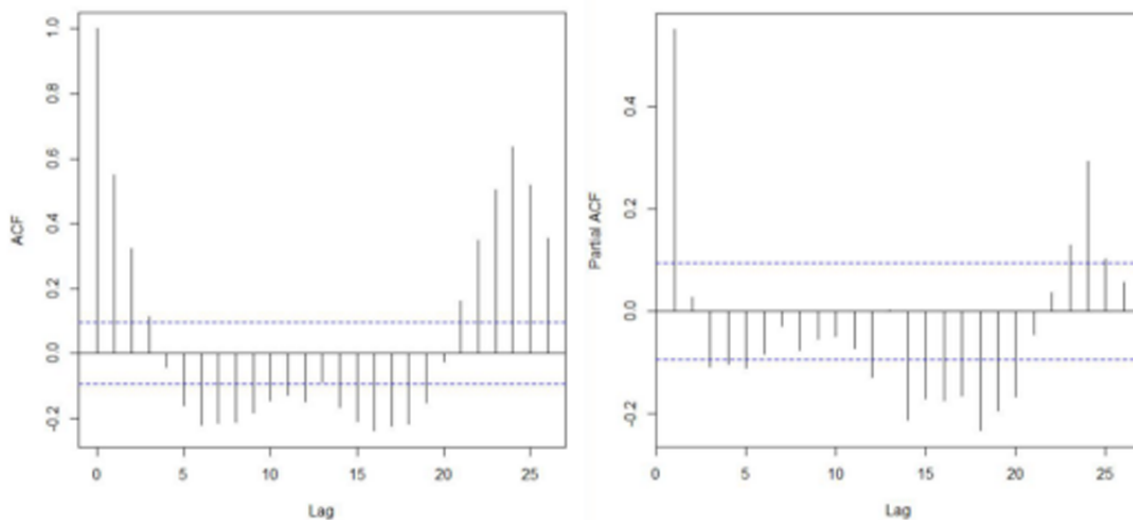


Figure 2. ACF,PACF graph

The $ARIMA(p, d, q) \times (P, D, Q)[S]$ model may have a lot of combination results. For these combination results, the optimal model can be determined by the AIC, SBC and AICC principle. The R model is used to determine the model as the optimal model according to the AIC criterion. We use the R modeling, determine the $ARIMA(1,0,0) \times (0,1,1)[24]$ as the optimal model according to the AIC criteria.

At the same time, the residuals of the model are tested, and the results show that the $p = 0.01$, $\chi^2 = 9.106$, which demonstrate that the series information of vehicles volume has been fully extracted, and the series model residuals are white noise series. The final model is identified as the optimal model[16].

Table 3. Combination Of Arima Seasonal Model

| Model | AIC | AICC | BIC |
|---------------------------|---------|---------|---------|
| ARIMA (1,0,0) (0,1,1)[24] | 5565.81 | 5565.87 | 5577.84 |
| ARIMA (1,0,0) (0,1,2)[24] | 5566.64 | 5566.74 | 5582.68 |
| ARIMA (1,0,0) (1,0,0)[24] | 6029.76 | 6029.86 | 6046.04 |
| ARIMA (1,0,1) (1,0,0)[24] | 6031.71 | 6031.85 | 6052.05 |
| ARIMA (1,0,2) | 6032.20 | 6032.4 | 6056.61 |

| Model | AIC | AICC | BIC |
|---------------------------|---------|---------|---------|
| (1,0,0)[24] | | | |
| ARIMA (1,0,3) (1,0,0)[24] | 6005.18 | 6005.44 | 6033.66 |
| ARIMA (1,0,0) (0,0,1)[24] | 6069.37 | 6069.47 | 6085.65 |
| ARIMA (1,1,0) (0,1,1)[24] | 5750.95 | 5751.01 | 5762.98 |

Predict Model use the $ARIMA(1,0,0) \times (0,0,1)[24]$ to predict the volume of the vehicle of the next 48 hour in the Kunming. The prediction results are shown in Table □, R^2 is 0.987, indicated that the model is highly fitted. The data fit graph is shown in the blue section of Figure 3. The model residual graph is shown in Figure 4.

Table 4. Arima Seasonal Model Fitting Degree Statistics

| Sufficiency | Average |
|---------------------|----------|
| Stationary R square | 0.987 |
| R Square | 0.990 |
| RMSE | 36.067 |
| MPAE | 9374.087 |

Forecasts from ARIMA(1,0,0)(0,1,1)[24]

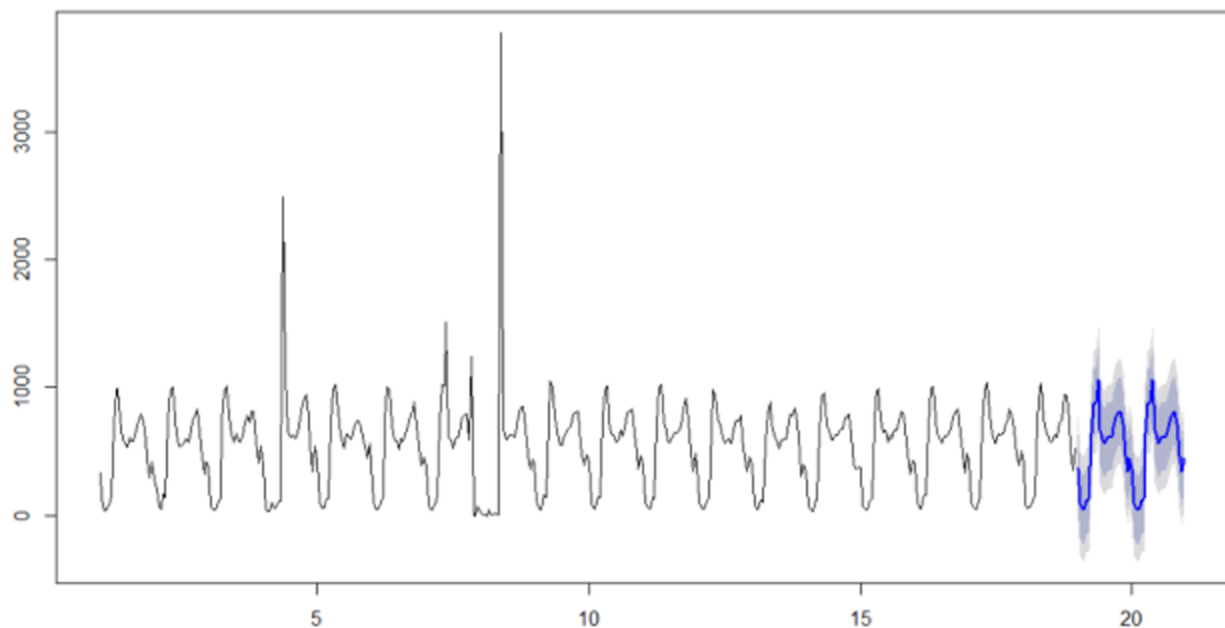


Figure 3. ARIMA Seasonal Model predict results graph

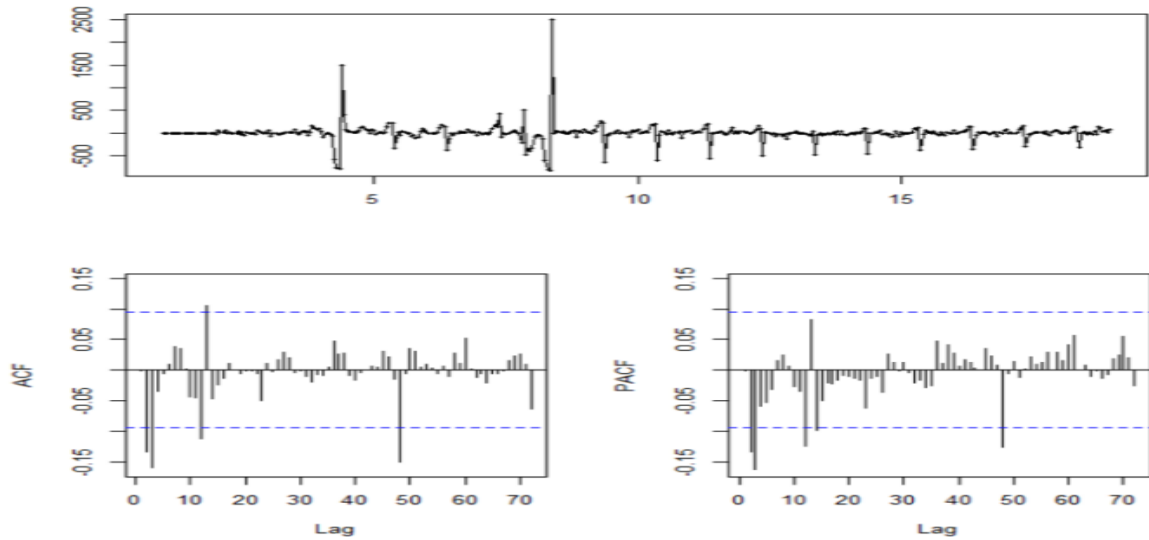


Figure 4. ARIMA Seasonal Model residual graph

4.2 Holt-Winters Experimental Results

The parameter alpha, beta and gamma of Holt-Winters Exponential smoothing model associated with the level, trend and the seasonal factors. The estimated value of alpha, beta and gamma are 0.0025, 0.2046 and 0.0734 respectively. The estimate of Alpha is 0.0025, indicating that the estimates at the current series time point balance

the historical observations and the most recent observations substantially. The estimate of beta is 0.2046, indicating that the slope of the estimated Holt-Winters trend is stable over the entire time series substantially. The estimate of gamma is 0.0734, which is very high, indicating that the seasonal part of the forecast at the current time point is based on historical observations, and the forecast results are shown in Figure 5.

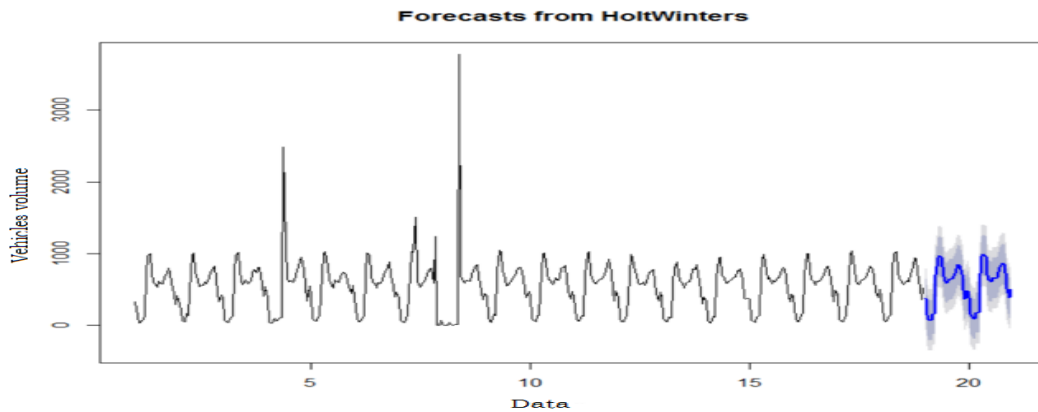


Figure 5. Holt-Winters forecast result graph

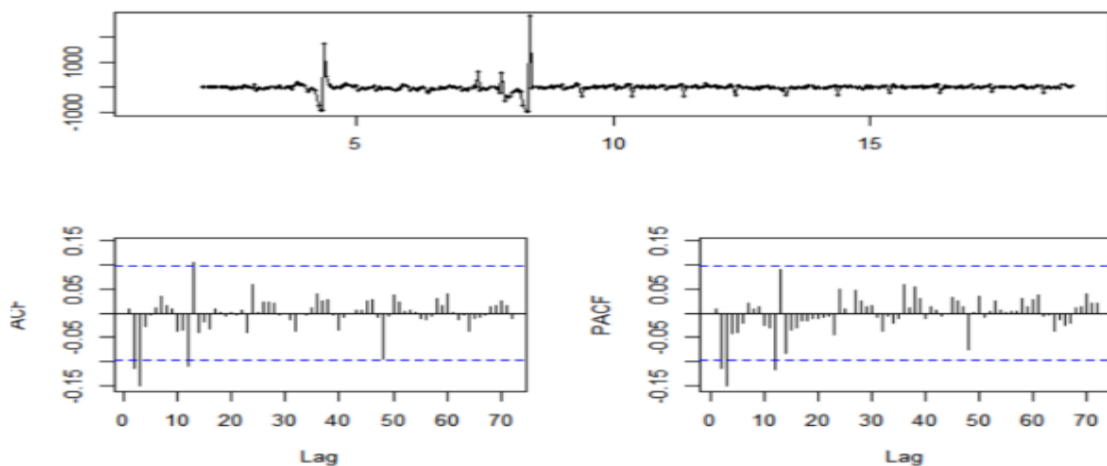


Figure 6. Holt-Winters residual Graph

Table 5. Holt-Winters Model Fitting Degree Statistics

| <i>Sufficiency</i> | <i>Average</i> |
|---------------------|----------------|
| Stationary R square | 0.745 |
| R Square | 0.673 |
| RMSE | 197.419 |
| MPAE | 119.188 |

Table 6 shows the experimental comparing result of the ARIMA and Holt-Winters. The ARIMA seasonal model has the higher predictability of vehicles volume data, and the predicted mean square error and mean error also lower than the Holt-Winters model.

Table 6. Model Fitting Degree Statistics

| <i>Model</i> | <i>Stationary R square</i> | <i>RMSE</i> | <i>AAD</i> |
|----------------------|----------------------------|-------------|------------|
| ARIMA seasonal model | 0.987 | 64.481 | 0.099 |
| Holt-Winters | 0.745 | 90.884 | 0.101 |

5 Conclusion

The ARIMA seasonal model and the Holt-Winters model are used to forecast vehicles volume of Kunming in Yunnan Province. The forecast value was compared with the actual value, we can conclude that the ARIMA seasonal model is used to predict the volume of the vehicle are better than Holt-Winters model, and the ARIMA seasonal model has a higher fitting degree and better forecasting error. Due to limited data, and there are many factors that affect the volume of the short-term vehicle, the forecast model has some errors. Other factors should be considered in the actual application.

Acknowledgment

This work has been supported by the Open Foundation of Key Laboratory in Software Engineering of Yunnan Province under Grant NO. 2017SE204.

References

- HongMeng Ji. Analysis on Passenger Transport Requirement and Passenger Volume Forecast of High Speed Railway[D].Beijing Jiaotong Universit,2015.
- Diongue A K, Diop A, Ndongo M. Seasonal fractional ARIMA with stable innovations[J]. Statistics & Probability Letters, 2008, 78(12):1404-1411.
- Wang Y, Wang J, Zhao G, et al. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China[J]. Energy Policy, 2012, 48(3):284-294.
- ZhenLong Wang. Time series analysis[M]. BeiJing: China Statistics Press, 2000.
- Kalekar P S. Time series Forecasting using Holt-Winters Exponential Smoothing[J]. Kanwal Rekhi School of Information Technology, 2004.
- ChingFu Chen, YuHern Chang, YuWei Chang. Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan[J]. Transportmetrica, 2009, 5(2):125-140.
- Montanari A, Rosso R, Taquq M S. A seasonal fractional ARIMA Model applied to the Nile River monthly flows at Aswan[J]. Water Resources Research, 2000, 36(5):1249-1259.
- Mohan S, Vedula S. Multiplicative seasonal Arima model for longterm forecasting of inflows[J]. Water Resources Management, 1995, 9(2):115-126.Article in a journal:
- Diongue A K, Diop A, Ndongo M. Seasonal fractional ARIMA with stable innovations[J]. Statistics & Probability Letters, 2008, 78(12):1404-1411.
- Aston J A D, Findley D F, Mcelroy T S, et al. New ARIMA Models for Seasonal Time Series and Their Application to Seasonal Adjustment and Forecasting[J]. Journal of Official Statistics, 2007(1).
- Kim B S, Hossein S Z, Choi G. Evaluation of temporal-spatial precipitation variability and prediction using seasonal ARIMA model in Mongolia[J]. KSCE Journal of Civil Engineering, 2011, 15(5):917-925.
- Chatfield C. The Holt-Winters Forecasting Procedure[J]. Journal of the Royal Statistical Society, 1978, 27(3):264-279.
- Chatfield C, Yar M. Holt-Winters Forecasting: Some Practical Issues[J]. Journal of the Royal Statistical Society, 1988, 37(2):129-140.
- Kalekar P S. Time series Forecasting using Holt-Winters Exponential Smoothing[J]. Kanwal Rekhi School of Information Technology, 2004.
- Chatfield C, Yar M. Prediction intervals for multiplicative Holt-Winters[J]. International Journal of Forecasting, 1991, 7(1):31-37.
- Roberts S A. A General Class of Holt-Winters Type Forecasting Models[M]. INFORMS, 1982.