SARIMA MODELLING APPROACH FOR RAILWAY PASSENGER FLOW FORECASTING

Miloš MILENKOVIĆ[1*], Libor ŠVADLENKA[2], Vlastimil MELICHAR[3],
Nebojša BOJOVIĆ[4], Zoran AVRAMOVIĆ[5]

[1, 4, 5]*Division for Management in Railway, Rolling stock and Traction,
The Faculty of Transport and Traffic Engineering, University of Belgrade, Serbia*
[2, 3]*Dept of Transport Management, Marketing and Logistics, Jan Perner Transport Faculty,
University of Pardubice, Czech Republic*

Abstract. In this paper, railway passenger flows are analyzed and a suitable modeling method proposed. Based on historical data composed from monthly passenger counts realized on Serbian railway network it is concluded that the time series has a strong autocorrelation of seasonal characteristics. In order to deal with seasonal periodicity, Seasonal AutoRegressive Integrated Moving Average (SARIMA) method is applied for fitting and forecasting the time series that spans over the January 2004 – June 2014 periods. Experimental results show good prediction performances. Therefore, developed SARIMA model can be considered for forecasting of monthly passenger flows on Serbian railways.

Keywords: railway, passenger service, time series, forecasting, SARIMA.

Introduction

Forecasting of future demand for transport service represent important element of success for a transport company. Forecasting also provides basic input for planning and control of functional areas including transport operations planning, marketing and finance.

There are two broad categories of forecasting approaches in transportation field: parametric and non-parametric (Smith *et al.* 2002; Vlahogianni *et al.* 2004; Wei, Chen 2012). The main difference between these two categories lies in functional dependence between independent variables and dependent variable. Within the non-parametric methods there are applications of neural networks (Dougherty 1995; Vlahogianni *et al.* 2004; Xie, Zhang 2006; Zhang, He 2007; Zhang, Ye 2008; Tsai *et al.* 2009; Alekseev, Seixas 2009), non-parametric regression (Smith *et al.* 2002; Williams *et al.* 1998; Clark 2003), Kalman filters (Whittaker *et al.* 1997; Stathopoulos, Karlaftis 2003; Saab, Zouein 2001; Wang *et al.* 2007) and Gaussian maximum likelihood (Tang *et al.* 2003). For example, Wei and Chen (2012) developed a hybrid forecasting approach based on a combination of Empirical Mode Decomposition (EMD) and Back-Propagation Neural (BPN) networks to predict short term passenger flow in metro systems. Xu *et al.* (2004) applied spatio-temporal data mining to forecast the railway passenger flow based on spatio. Approach is decomposed into two phases. In first phase, it forecasts the time sequence of the target object using statistical principles. In second phase, it figures out the spatial influence of neighbor objects using a neural network, and finally combines the two forecasting results using linear regression. Li *et al.* (2012) used Grey model for forecast annual outgoing passenger flows of a railway station. Authors analyzed monthly flucations in passenger flows and applied proportional coefficient method for prediction of passenger flows for each month.

In the traditional parametric techniques, historical average (Smith, Demetsky 1997; Stephanedes *et al.* 1981), smoothing techniques (Williams *et al.* 1998), and AutoRegressive Integrated Moving Average (ARIMA) (Hansen *et al.* 1999; Williams *et al.* 1998; Lee, Fambro 1999; Williams 2001; Kamarianakis, Prastacos 2005) have been applied to forecast transportation demand. For example, Grubb and Mason (2001) estimated a very long time series about air passenger traffic using Holt–Winters forecasting methods. They estimated the histor-

*Corresponding author. E-mails: *m.milenkovic@sf.bg.ac.rs, m.millenkovic@gmail.com*

Taylor & Francis
Taylor & Francis Group

ical growth using Holt–Winters decomposition and produced long lead-time forecasts. On the base of long data series, they evaluated the out-of-sample forecasting performance of this and other forecasting procedures and concluded that presented modification of Holt–Winter method greatly improves forecasting performance for long lead-times. Presented approach also considers an assessment of uncertainty in the predictions. By modified Holt–Winters procedure they varied the trend used for predictions and estimated the sensitivity of forecasts to assumptions about the future trend. Grosche *et al.* (2007) presented two gravity models for the estimation of air passenger flows between pairs of cities. Models are based on variables describing economic activity and geographical characteristics of city pairs. This means that developed models can be applied in case when new service is considered or historical data are not available.

Within the class of parametric methods, AutoRegressive Integrated Moving Average (ARIMA) has become one of the common parametric forecasting approaches since the 1970s (Wei, Chen 2012). In addition, with the characteristics of seasonality and trends in traffic data, some researches use seasonal ARIMA to predict traffic flow (Williams, Hoel 2003; Tan *et al.* 2009) and international air passenger flow (Faraway, Chatfield 1993; Lim, McAleer 2002; Chen *et al.* 2009).

As already mentioned, many railway planning processes including financial, asset, capital investment and service plan are predicted on the estimate of passenger demand to be handled within a given time period. Forecasts of traffic volume represent one of the most important outputs of a marketing plan and also the important input to the corporate plan. Current planning process in JSC 'Serbian Railways' lacks of any planning tool, so having that in mind, as well as the actual market share and increasing competition there is a need for better planning instruments in the company. Therefore, in this paper we present the first detailed application of Seasonal ARIMA (SARIMA) models in order to generate passenger traffic forecasts for the case of JSC 'Serbian Railways'. We use the data representing monthly passenger flows on all lines of Serbian railway network provided by Statistical Office of the Republic of Serbia (*http://webrzs. stat.gov.rs*). At the time of our analysis a time series of monthly passenger flows from January 2004 to June 2014 (126 months) was available.

Paper is organized as follows. Section 1 introduces ARIMA and SARIMA models and describes Box–Jenkins methodology for their selection. In Section 2, SARIMA modeling process has been applied for deriving an efficient model to assess rail passenger demand for a case of Serbian railways. Last section concludes this paper.

1. ARIMA and SARIMA models

Box *et al.* (2013) introduced the ARIMA method. This method now represents one of the most frequently used univariate time series modeling tools. ARIMA models are based on AutoRegressive Model (AR), the Moving Average Model (MA) and the combination of the AR and MA, the ARMA models (Suhartono 2011).

AR model includes lagged terms on the time series itself, and MA model includes lagged terms on the noise or residuals. The first requirement for ARIMA modeling is that the time series data to be modeled are either stationary or can be transformed into stationary. Therefore, the letter 'I' (Integrated) means that the first order difference is applied in order to stationarize given time series. First order differencing requires that there is a need to find between observations in two successive months. In case that there is a time series with trends, seasonal pattern and short time correlations a Seasonal ARIMA (SARIMA) model can be used. Besides three main components present in ARIMA model, there is a need for seasonal differencing in order to make a seasonal time series stationary. Therefore, a SARIMA model has four components:

– the non-seasonal and seasonal AR polynomial term of order *p* and *P*, respectively:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p; \tag{1}$$

$$\Phi_P(Bs) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}; \tag{2}$$

– the non-seasonal and seasonal Moving Average (MA) part of order *q* and *Q*, respectively:

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q; \tag{3}$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}; \tag{4}$$

– non-seasonal differencing operator, $(1-B)^d$, is the of order *d* used to eliminate polynomial trends;

– seasonal differencing operator, $(1-B^s)^D$, of order *D* used to eliminate seasonal patterns.

Parameters ϕ and θ are the ordinary ARMA coefficients, Φ and Θ are the seasonal ARMA coefficients, *B* is the backshift operator, whose effect on a time series Y_t can be summarized as $B^d Y_t = Y_{t-d}$.

Therefore, the generalized form of SARIMA$(p, d, q) \times (P, D, Q)_s$ model for a series Y_t can be written as (Box *et al.* 2013; Cryer, Chan 2008):

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \tag{5}$$

where: *s* is the length of the periodicity (seasonality) and ε_t is a white noise sequence.

Box–Jenkins method in essence involve three steps for fitting the ARIMA and SARIMA models. These are identification, estimation and validation of the model (Box *et al.* 2013; Prista *et al.* 2011). First step is generally based on an analysis of AutoCorrelation Function (ACF) and Partial ACF (PACF) and their comparison with theoretical profiles of these functions in AR, MA and ARMA processes. The output of the identification step is determined appropriate $(p, d, q) \times (P, D, Q)_S$ model structure. The model structure selected in previous step has to be fitted to the time series and its parameters to be estimated. This is the essence of the second step and it is done by using the conditional sum of squares or maximum likelihood method. Validation of selected model is
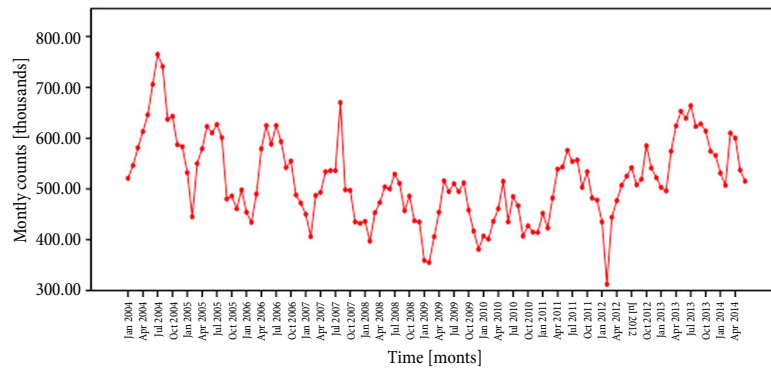
Figure 1. Time series of monthly rail passenger flows (in thousands) in Serbia (January 2004 – June 2014) – raw data

performed within the diagnostic checking by analysis of stationarity, invertibility as well as the presence of redundancy in model parameters. If a selected model fails in diagnostic check it is necessary to repeat the whole procedure again. After an appropriate model is found it can be used for forecasting purpose (Box *et al.* 2013; Prista *et al.* 2011).

Subjectivity is mostly present in model identification step. In the fact that this step is based essentially on graphical interpretations of ACF/PACF estimates. To cope with this subjectivity and to improve the determination the final orders of the ARMA processes there are a lot of model selection criteria proposed in literature (De Gooijer *et al.* 1985).

The most frequently used are information criteria such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and normalized version of BIC. These information criteria are designed to deal with the fit of the nonlinear models and to account for the number of the parameters in the model as well. They consist of the natural log of the Mean Squared Error (MSE) and a penalty for the number of parameters being estimated (Yaffee, McGee 2000):

$$\mathrm{AIC} = T\ln(\mathrm{MSE}) + 2k; \tag{6}$$

$$\mathrm{BIC} = T\ln(\mathrm{MSE}) + k\ln(T); \tag{7}$$

$$\mathrm{Normalized\ BIC} \;=\; \ln(\mathrm{MSE}) + k\frac{\ln(T)}{T}, \tag{8}$$

where: T is number of observations, k is the number of parameters in the model $k = p+q+P+Q+1$. Common procedure often involves estimating the largest model for which it is assumed to correctly capture the dynamics of a time series and then decreasing its size (dropping the lags) until the minimum value of AIC, BIC or Normalized BIC is reached. Presented methodology will be used in next chapter for analysis of railway passenger flows on Serbian railways.

2. SARIMA model for rail passenger flow forecasting: a case study of Serbian railways

In this chapter, we outline the practical steps, which need to be undertaken to use SARIMA time series models for passenger flows forecasting on Serbian railways. To test

alternative SARIMA models we obtained a time series of a total monthly number of passengers travelled by the railway from the Statistical Office of the Republic of Serbia (*http://webrzs.stat.gov.rs*). Data set covers the period from January 2004 to June 2014 (126 monthly values). The time series is presented on Figure 1. First 114 monthly observations are used for fitting purpose, whereas the remaining 12 observations served for verification the forecasting capability of selected SARIMA model.

According to the Box–Jenkins methodology, in SARIMA models there is an explicit consideration of trend and seasonal non-stationarities by the model structure. However, it is needed to analyze non-stationarities of variance before the model fitting. As it can be seen from the Figure 1, time series data $(x_t, t = 1,...,126)$ for rail passenger traffic in Serbia have strong seasonality pattern with a trend, which is not constant throughout the whole study period, rather slightly decreasing in period from 2004 to 2010 and increasing in period from 2011 to 2014, an increase in variance with the serious mean is identified from variance–mean plots (Figure 2).

Therefore, we log-transformed the data in order to stabilize the variance and then used these log transformed data set as input to the SARIMA analysis (Figure 3).

Different SARIMA models were applied to find the best fitting model. Among all candidate models, the most suitable predictive model for our case was found on the base of Box–Jenkins model and normalized BIC. This approach involved selection of the candidate model set, estimation of the model and determination of normalized BIC and diagnostic check.
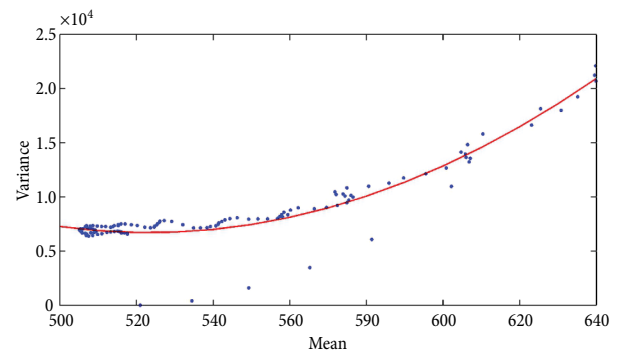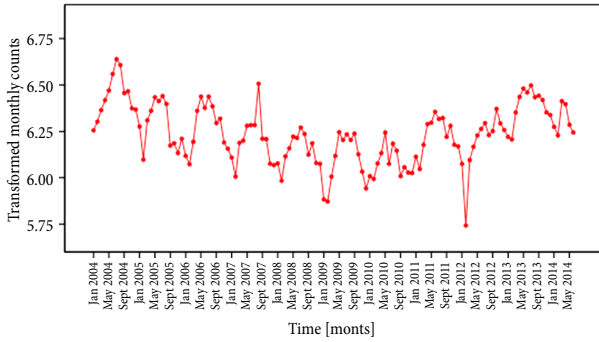

Figure 2. Mean–variance relationship

Figure 3. Time series of monthly rail passenger flows in Serbia (January 2004 – June 2014) – log transformed data

The set of alternative models is determined by analyzing the sample estimates of ACF and PACF, and defining the three major orders of the SARIMA models: d, D, and S. Lags 1, 2, 3, 4, 9, 10, 11, 12, 13, 14, 22, 23, 24, 25, 26, 29 and 30 had large autocorrelations with values 0.80, 0.64, 0.45, 0.28, 0.25, 0.36, 0.46, 0.51, 0.38, 0.27, 0.22, 0.31, 0.39, 0.29, 0.19, –0.17 and –0.26, respectively (Figure 4). According to these values, namely, the very sharp

decrease of autocorrelation values in first three lags; it is obvious that there exists a long-term trend. In order to eliminate it, the first lag difference term has to be included in the structure of the SARIMA model (d = 1). Annual lags and its multiplies also reported large autocorrelation values. Consequently, seasonal difference terms are included in models (S = 12, D = 1). Further justification of these conclusions is obvious from ACF and PACF plots of differenced time series (Figure 4). Based on this conclusions, we selected $\text{SARIMA}(p,1,q) \times (P,1,Q)_{12}$ as the basic structure of the alternative SARIMA models.

Among the statistical models, SARIMA $(0,1,0) \times (0,1,1)_{12}$ was the most appropriate with the lowest normalized BIC of 7.056 and a Mean Absolute Percent Error (MAPE) of 4.176 (Table 1). 70.8% of the variance of the time series was covered by this model (stationary R-squared). According to the Ljung–Box statistics it is proven that the selected model is correctly specified. Namely, the value of 0.542 shown here is not significant, so we can be confident in correct specification of the model. This model has the following equation:

$$\log Z_t = \log Z_{t-1} + \log Z_{t-12} - \log Z_{t-13} + \varepsilon_t + 0.966\varepsilon_{t-12}.$$
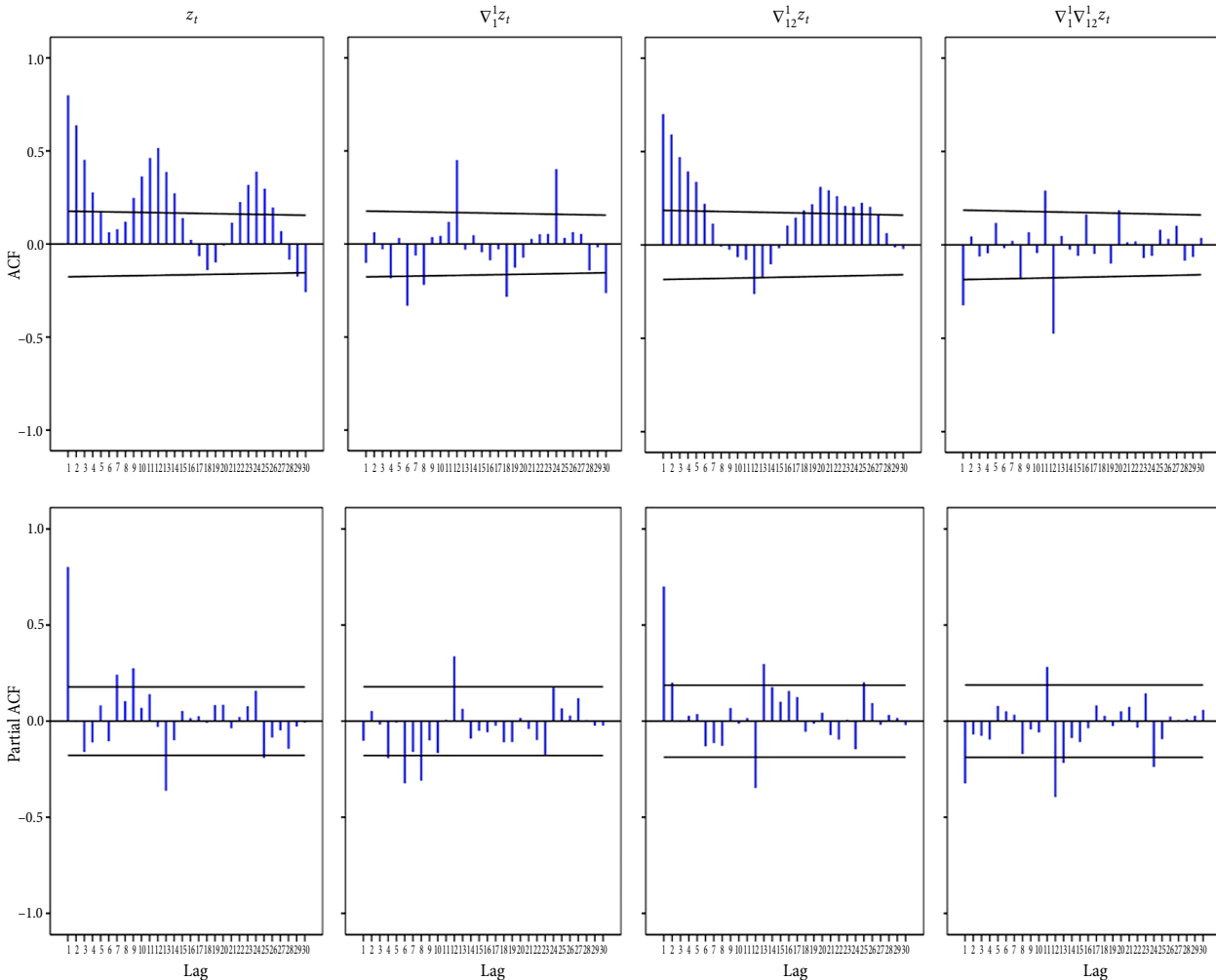
(9)

Figure 4. Sample ACF and PACF of the transformed rail passenger flow data. ACF/PACF plots for log transformed data (z_t, far left), lag-1 differenced series ($\nabla_1^1 z_t$), lag-12 differenced series ($\nabla_{12}^1 z_t$), and lag-1 and lag-12 differenced ($\nabla_1^1 \nabla_{12}^1 z_t$, far right) are displayed; the 95% confidence limits are represented by horizontal lines

Table 1. Normalized BIC, MAPE and stationary R-squared of SARIMA models

Models	MAPE	Normalized BIC	Stationary R-squared
SARIMA(0,1,1)×(0,1,1)$_{12}$	4.324	7.199	0.675
SARIMA(0,1,1)×(0,1,0)$_{12}$	5.452	7.443	0.545
SARIMA(0,1,0)×(0,1,1)$_{12}$	4.176	7.056	0.708
SARIMA(1,1,1)×(0,1,1)$_{12}$	4.303	7.233	0.679
SARIMA(1,1,1)×(1,1,1)$_{12}$	4.304	7.288	0.679

According to diagnostic checks SARIMA model appears as stationary and invertible without redundant parameters. According to the Ljung–Box Q-test and p-value (Ljung–Box $Q = 15.747$, p-value >0.05) residuals were white noise and therefore there is no significant auto-correlation between residuals at different lag times (Figure 5). Significance level of $\alpha = 0.05$ was used for all tests.

Twelve months of model forecasts are evaluated where the first month (July 2013) after the fitting part represent the origin of the forecast. Forecasts were obtained in the data original scale $\left(\hat{x}_l, l = 1, ..., 12\right)$. Performance of selected SARIMA model was evaluated by comparison of $l – step$ forecasts and monthly number of passengers travelled in period July 2013 – Jun 2014. In order to perform this comparison monthly forecast errors are computed and following accuracy criteria applied: Root Mean Square Error (RMSE), Mean Error (ME), Absolute Percent Error (APE$_l$) and MAPE. Performance of SARIMA model are compared also against the Seasonal Exponential Smoothing (SES) method (Table 2). For this purpose trial version of IBM SPSS (version 19) software is used.
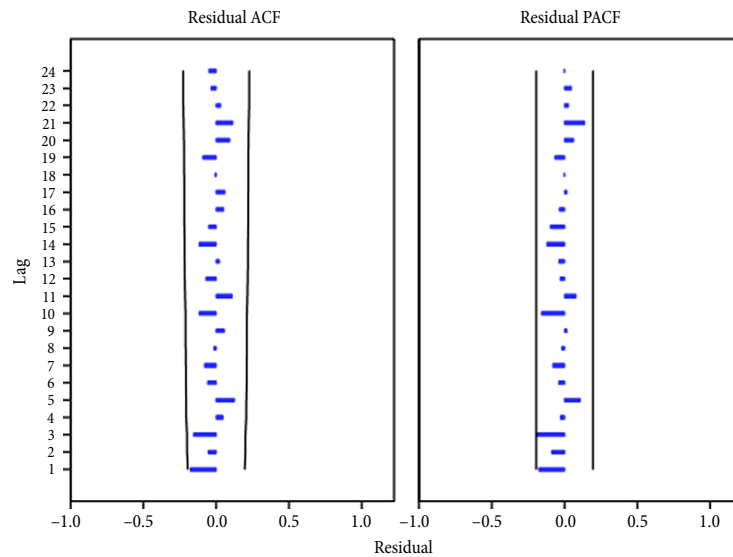

Figure 5. ACF and PACF of residuals

Table 2. Forecasts of railway passenger flows in thousands (July 2013 – June 2014): observed passenger counts (x_l), forecasted passenger counts ($\hat{x}_l$), monthly forecast errors (e_l) and monthly absolute percent error (APE$_l$) are displayed for Seasonal Exponential Smoothing (SES) model and SARIMA (SAR) model

Month, year	Step (l)	Obs(x_l)	Forecasts ($\hat{x}_l$)		Forecast errors (e_l)		APE$_l$	
			SES	SAR	SES	SAR	SES	SAR
July 2013	1	664	660.12	663.03	−3.88	−0.97	0.58	0.15
August 2013	2	623	655.84	641.02	32.84	18.02	5.27	2.89
September 2013	3	628	583.49	574.15	−44.51	−53.85	7.09	8.57
October 2013	4	614	598.45	598.03	−15.55	−15.97	2.53	2.60
November 2013	5	574	547.04	546.12	−26.96	−27.88	4.70	4.86
December 2013	6	566	540.73	539.27	−25.27	−26.73	4.46	4.72
January 2014	7	531	526.29	516.53	−4.71	−14.47	0.89	2.73
February 2014	8	507	485.6	488.51	−21.40	−18.49	4.22	3.65
March 2014	9	610	567.81	557.16	−42.19	−52.84	6.92	8.66
April 2014	10	600	613.22	601.82	13.22	1.82	2.20	0.30
May 2014	11	537	658.05	645.82	121.05	108.82	22.54	20.26
June 2014	12	515	649.67	649.48	134.67	134.48	26.15	26.11
Mean	1:12	580.75	590.52	585.08	9.77	4.33	7.30	7.13
Sum	1:12	6969	7086	7021	117	52		

Figure 6 displays observed values and SARIMA model fit and forecast values. Model gives slightly lower forecasts then observed values (September 2013 – March 2014), but pattern of model forecasts almost matched the one in observed passenger counts except for period from May 2014 to June 2014. The main reason for this is a slightly increasing trend for the last two years of the study period. RMSE during the prediction period (RMSE = 56.45) was 1.9 times the RMSE of the fitting period (RMSE = 29.69). Negative errors and low ME (ME = 4.32) registered in eight of twelve forecasts indicate on minority of underestimation in global terms. However, it has to be noted that this conclusion is valid for the case of random error, but this will not be the case if the error is systematic (system changes – transportation, activity or flow patterns). MAPE was 7.13% reflecting the slightly lower forecasts during the July 2013 and September 2013 – March 2014. The SARIMA model forecasts also registered better performances with respect to SES, resulting in 9.8% reduction in RMSE, 44.2% reduction in ME, and 9.7% reduction in MAPE.

Outliers are known to cause trouble in time series model identification, estimation and forecasts (Prista *et al.* 2011). Rail passenger flow data set presented four apparent outliers, which are incorporated into the model (Figure 7). They are of different kinds. These outliers are mainly additive that affects a single observation (August 2007, September 2009 and February 2012), but there is

also one transient outlier whose impact exponentially decays (June 2010). The main reason for existence of outliers at the beginning of year (like those in February 2012) lies in winter vacations and non-existence of daily and weekly school migrations. During the August 2007 significant flows appeared especially within the international passenger services. Higher volumes of passenger flows in September 2009 resulted by ending of summer vacations and beginning the activities at schools and faculties. June of 2010 in Serbia was very hot and this reflected on lower volumes of passenger flows on Serbian railways.

Conclusions

In this paper, we proposed a sophisticated mathematical model for generating rail passenger traffic forecasts for the case of JSC 'Serbian Railways'. Different SARIMA models were tested to select an appropriate.

The results of diagnostic check indicate that the SARIMA$(0,1,0)\times(0,1,1)_{12}$ is the most appropriate for modeling the rail passenger demand on Serbian railways.

This method can be a valuable decision support tool for company's marketing sector considering from one side, the current situation of rail passenger sector in Serbia (market share, revenues generated, number of passengers) and lack of any sophisticated planning tools, and from the other side, the accuracy of forecasting approach.
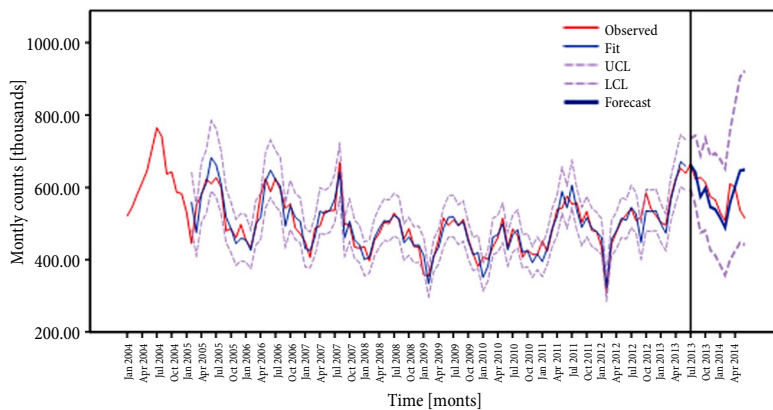

Figure 6. Results of SARIMA forecasting for the passenger demand on Serbian railways
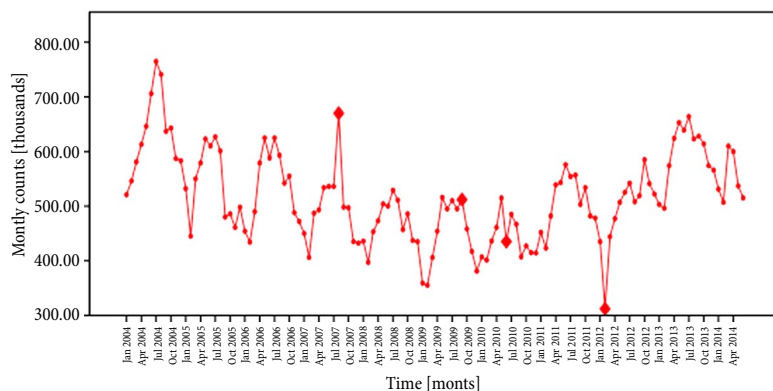

Figure 7. Outliers in time series of monthly rail passenger flows (in thousands) in Serbia (January 2004 – June 2014)

Future research will be concentrated on developing more efficient forecasting models, which will also use additional information from market surveys and demographic studies.

References

Alekseev, K. P. G.; Seixas, J. M. 2009. A multivariate neural forecasting modeling for air transport – preprocessed by decomposition: a Brazilian application, *Journal of Air Transport Management* 15(5): 212–216. http://dx.doi.org/10.1016/j.jairtraman.2008.08.008

Box, G. E. P.; Jenkins, G. M.; Reinsel, G. C. 2013. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Inc. 755 p. http://dx.doi.org/10.1002/9781118619193

Chatfield, C. 1993. Calculating interval forecasts, *Journal of Business & Economic Statistics* 11(2): 121–135. http://dx.doi.org/10.2307/1391361

Chen, C.-F.; Chang, Y.-H.; Chang, Y.-W. 2009. Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan, *Transportmetrica* 5(2): 125–140. http://dx.doi.org/10.1080/18128600802591210

Clark, S. 2003. Traffic prediction using multivariate nonparametric regression, *Journal of Transportation Engineering* 129(2): 161–168. http://dx.doi.org/10.1061/(ASCE)0733-947X(2003)129:2(161)

Cryer, J. D.; Chan, K.-S. 2008. *Time Series Analysis: with Applications in R*. Springer. 491 p. http://dx.doi.org/10.1007/978-0-387-75959-3

De Gooijer, J. G.; Abraham, B.; Gould, A.; Robinson, L. 1985. Methods for determining the order of an autoregressive-moving average process: a survey, *International Statistical Review / Revue Internationale de Statistique* 53(3): 301–329. http://dx.doi.org/10.2307/1402894

Dougherty, M. 1995. A review of neural networks applied to transport, *Transportation Research Part C: Emerging Technologies* 3(4): 247–260. http://dx.doi.org/10.1016/0968-090X(95)00009-8

Faraway, J.; Chatfield, C. 1998. Time series forecasting with neural networks: a comparative study using the air line data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47(2): 231–250. http://dx.doi.org/10.1111/1467-9876.00109

Grosche, T.; Rothlauf, F.; Heinzl, A. 2007. Gravity models for airline passenger volume estimation, *Journal of Air Transport Management* 13(4): 175–183. http://dx.doi.org/10.1016/j.jairtraman.2007.02.001

Grubb, H.; Mason, A. 2001. Long lead-time forecasting of UK air passengers by Holt–Winters methods with damped trend, *International Journal of Forecasting* 17(1): 71–82. http://dx.doi.org/10.1016/S0169-2070(00)00053-4

Hansen, J. V.; McDonald, J. B.; Nelson, R. D. 1999. Time series prediction with genetic-algorithm designed neural networks: an empirical comparison with modern statistical models, *Computational Intelligence* 15(3): 171–184. http://dx.doi.org/10.1111/0824-7935.00090

Kamarianakis, Y.; Prastacos, P. 2005. Space–time modeling of traffic flow, *Computers & Geosciences* 31(2): 119–133. http://dx.doi.org/10.1016/j.cageo.2004.05.012

Lee, S.; Fambro, D. 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting, *Transportation Research Record* 1678: 179–188. http://dx.doi.org/10.3141/1678-22

Li, H. J.; Zhang, Y.-Z.; Zhu, C.-F. 2012. Forecasting of railway passenger flow based on Grey model and monthly proportional coefficient, in *2012 IEEE Symposium on Robotics and Applications (ISRA)*, 3–5 June 2012, Kuala Lumpur, 23–26. http://dx.doi.org/10.1109/ISRA.2012.6219110

Lim, C.; McAleer, M. 2002. Time series forecasts of international travel demand for Australia, *Tourism Management* 23(4): 389–396. http://dx.doi.org/10.1016/S0261-5177(01)00098-X

Prista, N.; Diawara, N.; Costa, M. J.; Jones, C. 2011. Use of SARIMA models to assess data-poor fisheries: a case study with a sciaenid fishery off Portugal, *Fishery Bulletin* 109(2): 170–185.

Saab, S. S.; Zouein, P. P. 2001. Forecasting passenger load for a fixed planning horizon, *Journal of Air Transport Management* 7(6): 361–372. http://dx.doi.org/10.1016/S0969-6997(01)00030-8

Smith, B.; Demetsky, M. 1997. Traffic flow forecasting: comparison of modeling approaches, *Journal of Transportation Engineering* 123(4): 261–266. http://dx.doi.org/10.1061/(ASCE)0733-947X(1997)123:4(261)

Smith, B. L.; Williams, B. M.; Oswald, R. K. 2002. Comparison of parametric and nonparametric models for traffic flow forecasting, *Transportation Research Part C: Emerging Technologies* 10(4): 303–321. http://dx.doi.org/10.1016/S0968-090X(02)00009-8

Stathopoulos, A.; Karlaftis, M. G. 2003. A multivariate state space approach for urban traffic flow modeling and prediction, *Transportation Research Part C: Emerging Technologies* 11(2): 121–135. http://dx.doi.org/10.1016/S0968-090X(03)00004-4

Stephanedes, Y. J.; Michalopoulos, P. G.; Plum, R. A. 1981. Improved estimation of traffic flow for real-time control (discussion and closure), *Transportation Research Record* 795: 28–39.

Suhartono. 2011. Time series forecasting by using seasonal autoregressive integrated moving average: subset, multiplicative or additive model, *Journal of Mathematics and Statistics* 7(1): 20–27. http://dx.doi.org/10.3844/jmssp.2011.20.27

Tan, M.-C.; Wong, S. C.; Xu, J.-M.; Guan, Z.-R.; Zhang, P. 2009. An aggregation approach to short-term traffic flow prediction, *IEEE Transactions on Intelligent Transportation Systems* 10(1): 60–69. http://dx.doi.org/10.1109/TITS.2008.2011693

Tang, Y. F.; Lam, W. H. K.; Ng, P. L. P. 2003. Comparison of four modeling techniques for short-term AADT forecasting in Hong Kong, *Journal of Transportation Engineering* 129(3): 271–277. http://dx.doi.org/10.1061/(ASCE)0733-947X(2003)129:3(271)

Tsai, T.-H.; Lee C.-K.; Wei, C.-H. 2009. Neural network based temporal feature models for short-term railway passenger demand forecasting, *Expert Systems with Applications* 36(2): 3728–3736. http://dx.doi.org/10.1016/j.eswa.2008.02.071

Vlahogianni, E. I.; Golias, J. C.; Karlaftis, M. G. 2004. Short-term traffic forecasting: overview of objectives and methods, *Transport Reviews* 24(5): 533–557. http://dx.doi.org/10.1080/0144164042000195072

Wang, Y.; Papageorgiou, M.; Messmer, A. 2007. Real-time freeway traffic state estimation based on extended Kalman filter: a case study, *Transportation Science* 41(2): 167–181. http://dx.doi.org/10.1287/trsc.1070.0194

Wei, Y.; Chen, M.-C. 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks, *Transportation Research Part C: Emerging Technologies* 21(1): 148–162. http://dx.doi.org/10.1016/j.trc.2011.06.009

Whittaker, J.; Garside, S.; Lindvel, K. 1997. Tracking and predicting a network traffic process, *International Journal of Forecasting* 13(1): 51–61.
http://dx.doi.org/10.1016/S0169-2070(96)00700-5

Williams, B. M. 2001. Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling, *Transportation Research Record: Journal of the Transportation Research Board* 1776: 194–200. http://dx.doi.org/10.3141/1776-25

Williams, B. M.; Durvasula, P. K.; Brown, D. E. 1998. Urban Freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models, *Transportation Research Record* 1644: 132–141. http://dx.doi.org/10.3141/1644-14

Williams, B. M.; Hoel, L. A. 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results, *Journal of Transportation Engineering* 129(6): 664–672.
http://dx.doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)

Xie, Y.; Zhang, Y. 2006. A wavelet network model for short-term traffic volume forecasting, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 10(3): 141–150. http://dx.doi.org/10.1080/15472450600798551

Xu, W.; Qin, Y.; Huang, H. 2004. A new method of railway passenger flow forecasting based on spatio-temporal data mining, in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems*, 3–6 October 2004, Washington, USA, 402–405.
http://dx.doi.org/10.1109/ITSC.2004.1398932

Yaffee, R. A.; McGee, M. 2000. *An Introduction to Time Series Analysis and Forecasting: with Applications of SAS and SPSS*. Academic Press. 528 p.

Zhang, X.-L.; He, G.-G. 2007. Forecasting approach for short-term traffic flow based on principal component analysis and combined neural network, *Systems Engineering – Theory & Practice* 27(8): 167–171.
http://dx.doi.org/10.1016/S1874-8651(08)60052-6

Zhang, Y.; Ye, Z. 2008. Short-term traffic flow forecasting using fuzzy logic system methods, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 12(3): 102–112. http://dx.doi.org/10.1080/15472450802262281