International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 11 Issue 2 || Mar. - Apr., 2023 || PP 03-07

Air Traffic Prediction Using Machine Learning

Mounika, Panjala and Bhatracharyulu, N.Ch.
mounikapanjala123@gmail.com and dwarakbhat@osmania.ac.in
Department of Statistics, University College of Science,
Osmania University, Hyderabad, Telangana, India.

Abstract
This study aims to fit an appropriate model to forecast the number of domestic air passengers travelled through Spice Jet during the period 2009 to 2019. The methods used to forecast are, Holt-Winter, autoregressive integrated moving average (ARIMA), seasonal ARIMA and feed forward neural networks. The prediction performance was evaluated using various error metrics.

Keywords: *Spice Jet, Holt-Winter, Seasonal ARIMA, MLP, Error metrics.*

Date of Submission: 27-04-2023 Date of Acceptance: 15-05-2023

I. INTRODUCTION

The airline industry is one of the major services industries that considerably add to the wealth of developed and developing nations. It is one of the most vital sectors in the economic progress of a nation. It serves a key role in transporting people or products from one location to another, especially during short and distant transportation. Several researchers are working on prediction of air traffic data using various time series models.

This study is focused on the air traffic of Spice Jet. The number of passengers travelled in a month through domestic Spice Jet during the period 2009 to 2019 (sample size 132) was collected from website of civil aviation. The average number of passengers travelled through spice jet is 9,93,572 (around), with a standard deviation 3,88,715.86. the data is positively skewed with skewness value 0.6789. and the data is platy kurtic with kurtosis 0.2125. The range of passengers travelled in a month is lies between 3,78,758 and 21,44,156. The instability in the data was examined using Cuddy Della Valle Index (17.91). The Coefficient of variation in the data is 0.3912%. The data plot is presented in Fig.1.

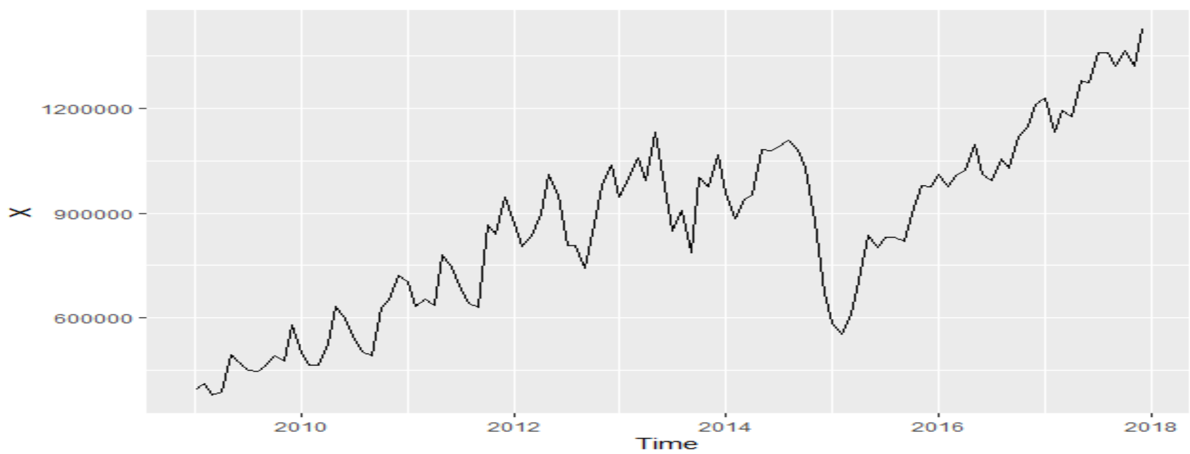

Fig-1: Data plot of passengers travelled through Spice Jet

II. TIME SERIES MACHINE LEARNING TECHNIQUES

To forecast the future values of the time series Holts-winter triple exponential smoothing technique, ARIMA and seasonal ARIMA and multilayer perceptron methods were used due to seasonality and trend in the data.

Holt-Winters' method is a triple exponential smoothing model is appropriate when trend and seasonality, which is a complicating factor present in the data. The additive and multiplicative models were preferred based on the seasonal variations are nearly constant throughout the series, or changing proportional to the level of the series. Let L_t denotes level of series, b_t represents trend, S_t denotes the seasonal component in the time series data Y_t and F_{t+m} denotes forecast of 'm' periods into the future, then the holts winter additive model is, $F_{t+m} = L_t + b_{tm} + S_{t-s+m}$ where $L_t = \alpha\,(Y_t - S_{t-s}) + (1-\alpha)\,(L_{t-1} + b_{t-1})$; $b_t = \beta(L_t - L_{t-1}) + (1-\beta)\,b_{t-1}$ $S_t = \gamma\,(Y_t - L_t) + (1-\gamma)\,S_{t-s}$ and where α, β, γ are level, trend, and seasonality smoothing constants and are lies between 0 and 1. The holts winter multiplicative model is, $F_{t+m} = (L_t + b_{tm})\,S_{t-s+m}$ where $L_t = \alpha\,(Y_t\,S_{t-s}) + (1-\alpha)\,(L_{t-1} + b_{t-1})$; $b_t = \beta(L_t - L_{t-1}) + (1-\beta)\,b_{t-1}$ $S_t = \gamma\,(Y_t\,L_t) + (1-\gamma)\,S_{t-s}$ and where α, β, γ are level, trend, and seasonality smoothing constants and are lies between 0 and 1.

The autoregressive integrated moving average (ARIMA) is an advanced forecasting technique introduced by Box and Jenkins (1977). The ARIMA (p, d, q) model is $\phi\,(B)\,(1-B)^d X_t = \theta(B)\,a_t$; where $\phi\,(B) = 1 - \phi_1 B - \cdots - \phi_p B_p$; and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q Bq$; Where p is lagged dependent variables in AR and q lagged error terms and with an order of differencing d. Seasonal ARIMA is an additive model to ARIMA when seasonality is observed in the data. It is denoted by ARIMA $(p, d, q)\,(P, D, Q)_s$. The model is: $\phi\,(B)\,\Phi(B_S)\,(1-B)^d\,(1-B_S)^D Xt = \theta(B)\,\Theta(B_S) + a_t$ where: a_t be the white noise with mean zero and variance σ^2; 'd' be the number of regular differences; $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B_p$ and $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_P B_P$; $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q Bq$ and $\Theta(B) = 1 + \Theta_1 B + \cdots + \Theta_Q B_Q$.

A multi-layer Perceptron is a feed forward, supervised, multi-layer neural network learning technique. The neurons are organized in three layers: input layer, hidden or intermediate layer and output layer. Each of the neuron is used to identify small linearly separable sections of the inputs. Outputs of the neurons are combined into next layer neurons and so on to produce the final output.

III. AIR TRAFFIC PREDICTION MODELS

Estimated parameters of the time series models used for fitting the Spice Jet data and their residual plots with lags are presented below. are:

Residual plots of time series models

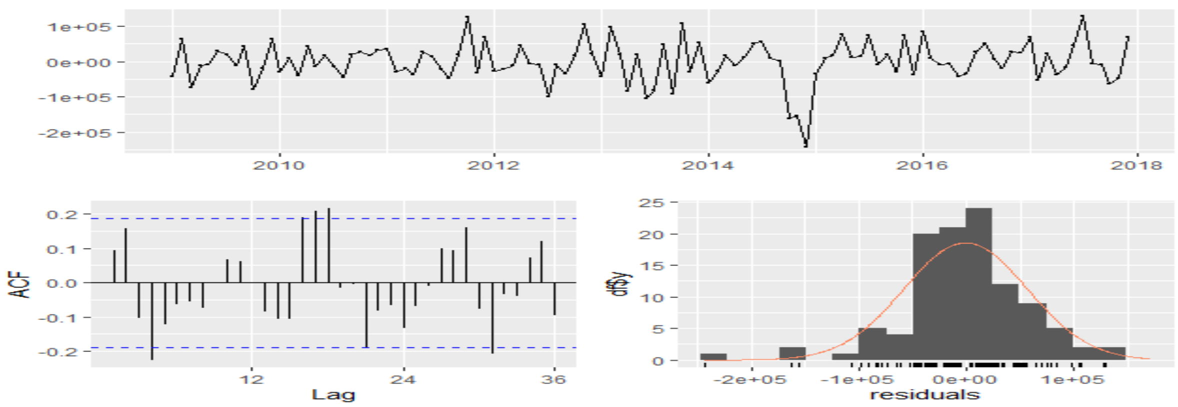

Fig. 3.1(a): Holts-Winter model: (Additive) α= 0.9999; β= 0.0001 and γ= 0.0001;

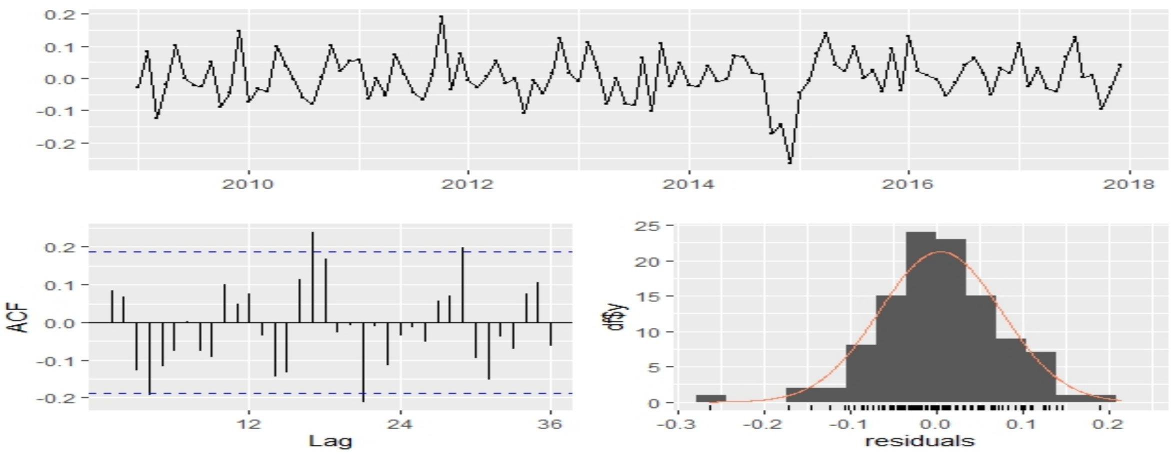

Fig. 3.1(b): Holts-Winter model: (Multiplicative) α= 0.9975; β= 0.0005 and γ= 0.0023;

Residual plots of time series models

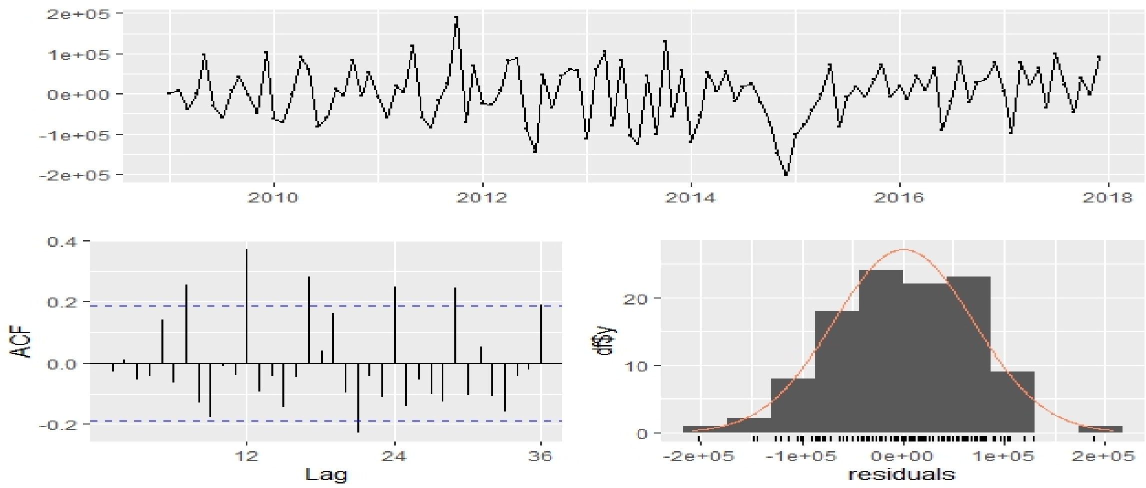

Fig 3.1(c): ARIMA (0,1,4) model: $(1-B)Y_t = 1-0.0319B - 0.1434B_2 + 0.2151B_3 + 0.3868B_4$

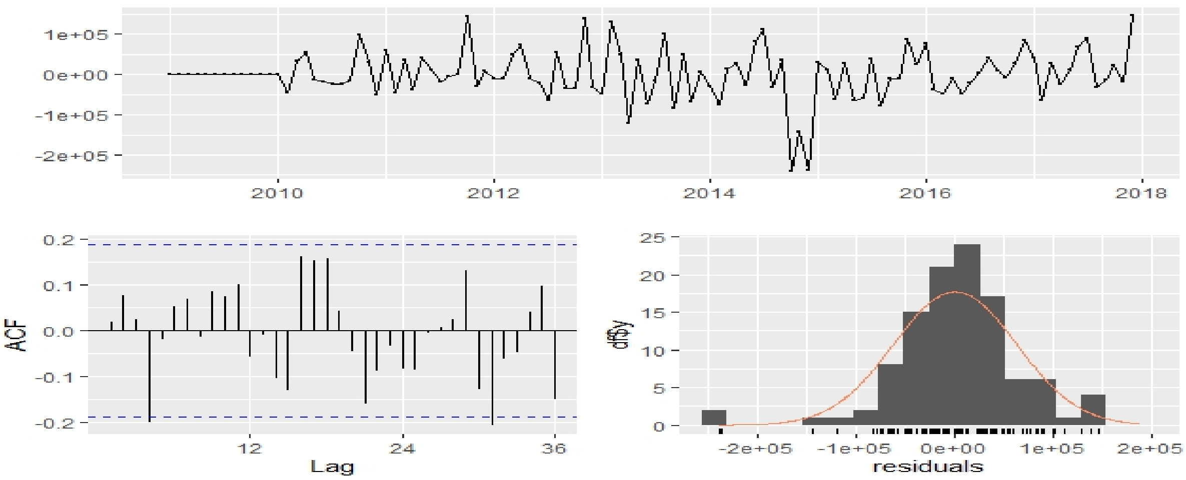

Fig. 3.1(d): ARIMA (3,0,0) (2,1,0) [12]

$(1-1.022B-0.0997B_2-0.2446B_3)(1+0.4881B+0.2454B_2)(1-B)(1-B_s)Y_t$.

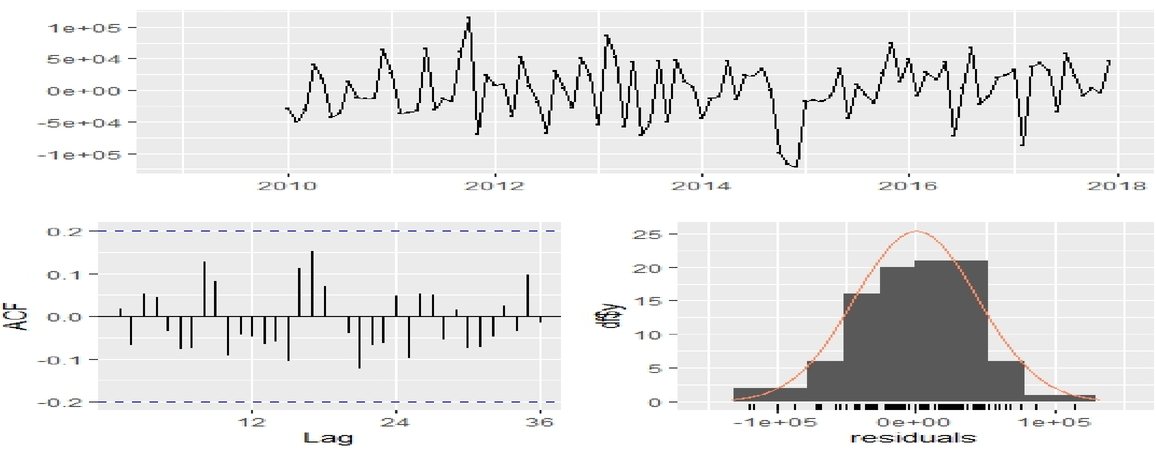

Fig. 3.1(e): Multi-layer Perceptron Neural Network model with variance 2.09E+09

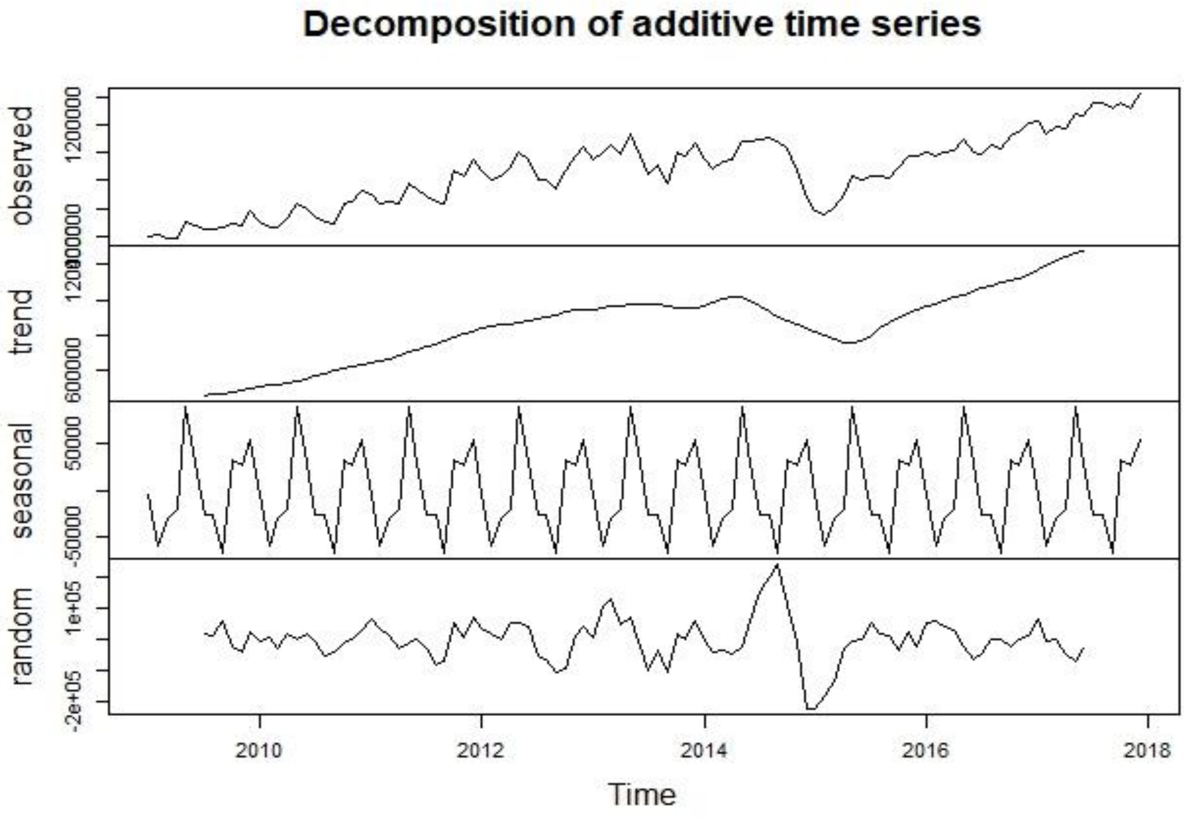

Fig 3.2: Decomposition of seasonality and trend from the observed data.

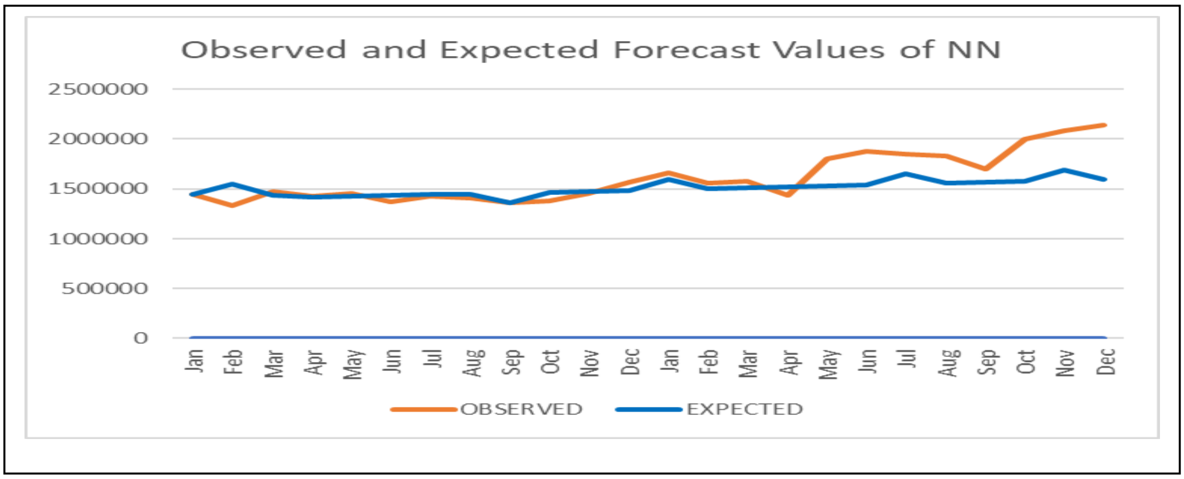

Fig. 3.3: comparison Observed and predicted values

Table 3.1 Comparison of prediction errors and information criteria

Forecasted Model	RMSE	MAPE	MASE	AIC	BIC
Holts-Winter Model Additive	57130.8	5.3404	0.2381	2905.54	2951.14
Holts-winter Model Multiplicative	62276.8	5.36191	0.25103	2908.61	2956.88
ARIMA Model (1,1,0)	68834.6	6.84311	0.30906	2701.46	2717.5
ARIMA Model (3,0,0) (2,1,0) [12]	62314.6	4.98358	0.24446	2422.97	2440.92
NNAR (12,1,2) [12]	44142.4	4.07533	0.19555	-	-

Remarks:
1. The instability in the data was examined using Cuddy Della Valle Index (17.91) It is indicating medium
2. Stationarity in the data was tested through Augmented Dickey-Fuller test and found non-stationarity. The Dickey-Filler test statistic is -1.4796 with lag order= 4, with p= 0.7925.
3. The seasonality and trend significances can be observed from Fig 3.2.
4. The significance in the estimated parameters of AR1, AR2, AR3, SAR1, SAR2 and Drift was examined using Lagranger Multiplier test with Z-values as 10.0630, 0.6643, -2.4138, -4.6226, -2.4684 and 3.1154 with p = 2.2e-16, 0.506505, 0.015785, 3.79e-06, 0.013571, 0.001837.
5. The White Noise process 'a_t' is assumed to be i.i.d. Normal (0, σ^2) can be verified from the Time sequence plot of residuals. The mean of the residuals are surrounded to zero with more fluctuations in all methods.
6. The least value in RMSE, MAPE, MASE out of all the models fitted is corresponding to neural Network model**.**
7. Based on the AIC and BIC values in the functional models seasonal ARIMA is most fitted.
8. The autocorrelation in residuals of the SARIMA (3,0,0) (2,1,0) [12] was tested using Ljung-Box test (Q* = 28.044 with 17 d.f. and 5 model parameters and total 22 lags) Its significance p =0.04443. There is no significant correlation in the residual series, there are few spikes in the graphs indicating a correlation between the lags.
9. The histogram suggests that the residuals are how closed to Normal.

REFERENCES

[1]. Albayrak, M. B. K., Özcan, İ. Ç., Can, R., & Dobruszkes, F. (2020): The determinants of air passenger traffic at Turkish airports, Journal of Air Transport Management, vol 86, pp 101-108.
[2]. Chen, J., Chen, L., & Sun, D. (2017): Air traffic flow management under uncertainty using chance-constrained optimization, Transportation Research Part B: Methodological, vol 102, pp 124-141.
[3]. Gelhausen, M. C., Berster, P., & Wilken, D. (2018): A new direct demand model of long-term forecasting air passengers and air transport movements at German airports, Journal of Air Transport Management, vol 71, pp 140-152.
[4]. Jin, F., Li, Y., Sun, S., & Li, H. (2020): Forecasting air passenger demand with a new hybrid ensemble approach, Journal of Air Transport Management, vol 83, pp 101-104.
[5]. Lin, Y., Zhang, J. W., & Liu, H. (2019): Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation, Aerospace Science and Technology, vol 93, pp 105-113.
[6]. Sun, S., Lu, H., Tsui, K. L., & Wang, S. (2019): Nonlinear vector auto-regression neural network for forecasting air passenger flow, Journal of Air Transport Management, vol 78, pp 54-62.