

# CHARACTERIZATION OF AIRPORT DELAY

**Bianka Karoly<sup>1</sup>, Balazs Saghi<sup>2</sup>**

<sup>1</sup> HungaroControl Ltd. H-1185 Budapest, Igló utca 33-35, Hungary

<sup>1,2</sup> Budapest University of Technology and Economics, Faculty of Transport Engineering and Vehicle Engineering, Department of Control for Transportation and Vehicle Systems, H-1111, Muegyetem rkp 3, Hungary

Received 4 June 2019; accepted 27 September 2019

**Abstract:** Scheduling and coordinating the departures and arrivals at the airport can be difficult if there is a delay that has to be taken into account, but relying on a statistical model that bases on historical data could result in better prediction for the future. Therefore, in this study the authors investigate the nature of the airport delays with statistical tools. First the airport delay is defined and the hypothesis was formed. In this paper authors investigated if airport delay follows a normal distribution. Moreover detailed description of used statistical tools are mentioned. In third paragraph the results are examined briefly. As conclusion authors found that generally airport delay not followed normal distribution.

**Keywords:** airport delay, statistical analysis, normal distribution.

## 1. Introduction

As the air traffic demand, and as a consequence the air traffic, increases, there is constant need to decrease the flight delay which. The airports are the hotspot of the whole air traffic chain. Therefore measuring and modelling the delay that arises at the airport could help in reducing and illuminating the delays for the whole chain. Due to that reason the authors feel need to investigate statistically the cases (Kling *et al.*, 2017; Madácsi, 2015).

The authors' assumption is that the deviation of the actual and the planned departing and arriving time at the airport follows Normal distribution. To carry out a statistical assessment, data were collected about typical traffic at Budapest Liszt Ferenc International Airport.

## 2. Methodology

The study was conducted on average daily traffic during a summer season of Budapest Airport Liszt Ferenc International Airport. The elementary event of the current study is the difference between the scheduled and the actual departure or arrival time, that we call as delay:

$$t_{\text{delay}} = T_{\text{actual}}^{\text{departure}} - T_{\text{scheduled}}^{\text{departure}} \quad (1)$$

$$t_{\text{delay}} = T_{\text{actual}}^{\text{arrival}} - T_{\text{scheduled}}^{\text{arrival}} \quad (2)$$

It is important to note that, this delay value in our context can be negative: if the aircraft arrives or departs at the airport before the scheduled time. The measured elementary events considered discrete distributed variable (Serhan *et al.*, 2018; Čokorilo *et al.*, 2010).

<sup>1</sup> Corresponding author: [bianka.karoly@hungarocontrol.hu](mailto:bianka.karoly@hungarocontrol.hu)

## 2.1. Boxplot Analysis

Boxplots are a standardized way of displaying the distribution of data based on a five number summary, which are:

- minimum;
- first quartile ( $Q_1$ );
- median;
- third quartile ( $Q_3$ ); and
- maximum.

The Boxplot displays the full range of variation (from min to max), the likely range of variation interquartile range (IQR) which is the range of the variation from the  $Q_1$  to the  $Q_3$ . The median presents as a typical value of the distribution. Boxplot also visualize the

surprisingly high or surprisingly low values called outliers. Two types of outliers can be distinguished:

Suspected outliers that are either  $1,5 \cdot IQR$  or more above the  $Q_3$  or  $1,5 \cdot IQR$  or more below the  $Q_1$ :

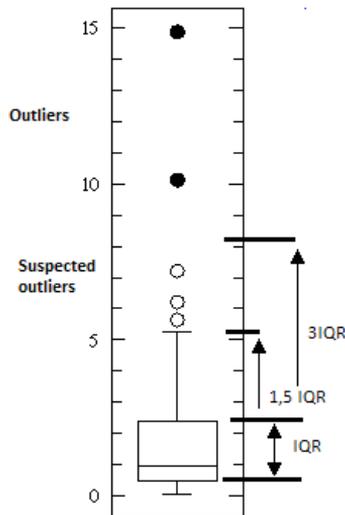
$$x_i < Q_1 - 1,5IQR \quad (3)$$

$$x_i > Q_3 + 1,5IQR \quad (4)$$

Outliers are either  $3 \cdot IQR$  or more above the  $Q_3$  or  $3 \cdot IQR$  or more below the  $Q_1$ :

$$x_i < Q_1 - 3IQR \quad (5)$$

$$x_i > Q_3 + 3IQR \quad (6)$$



**Fig. 1.**  
*Interpretation of Boxplot Diagram*  
Source: own analysis

Suspected outliers means in our context that the average delay on that day is significantly larger or smaller than the other deviation on the same day. Therefore, these could be

values that are associated with abnormal, one-time events or special causes. These could have a great influence on the final results of the analysis. Therefore, it was

important not to take them into consideration and therefore the outliers were removed from the sample (Čokorilo and Dell’Acqua, 2013).

The authors assumption that the filtered dataset follows Normal distribution which function is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{7}$$

### 2.2. Graphical Test for Normality

To justify your hypothesis at first we applied graphical test. Graphical tests are for comparing two probability distributions by plotting their shapes against each other.

With SPSS software the Q-Q (quantile-quantile) normality test was chosen that fits well for normality test purposes. The purpose of Q-Q plots is to find out if two sets of data come from the same distribution. An identity function is plotted on the Q-Q plot; if the two data sets come from a common distribution, the points will fall on that reference line. Comparing them against each other the subjective decision can be made to apply or reject the initial hypothesis (Ghasemi and Zahediasl, 2012; Ford, 2018).

### 2.3. Chi Square Test for Normality

To complete the subjective method with an objective one, Chi-square ( $\chi^2$ ) goodness of fit test was also conducted. Chi-Square goodness of fit test is a non-parametric test

that is used to compare the observed sample distribution with the expected probability distribution. Chi-Square goodness of fit test determines how well theoretical distribution (Nominal in this case) fits the empirical distribution.

For the calculation of the value of Chi-Square goodness of fit test, the following formula was used:

$$\chi^2_{calc} = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \tag{8}$$

where:

$\chi^2_{calc}$  = Calculated Chi-Square goodness of fit test;

O= observed value;

E= expected value.

The null hypothesis ( $H_0$ ) of the study assumption was that there is no significant difference between the observed distribution of the samples and the expected Normal distribution. While, the alternative hypothesis ( $H_1$ ) assumes that there is a significant difference.

In Chi-Square goodness of fit test, the degree of freedom depends on the distribution of the sample. The degrees of freedom is the number of values in the calculation of a statistic that are free to vary. Table 1 shows the different distributions and the associated degree of freedom (Čokorilo and Dell’Acqua, 2013; Miroslavljević *et al.*, 2011; Sipos, 2017).

**Table 1**  
*Distributions and the Associated Degree of Freedom*

Type of Distribution	No of Constraints	Degree of Freedom
Binominal distribution	1	n-1
Poisson distribution	2	n-2
Normal distribution	3	n-3

The calculated value of Chi-Square goodness of fit test was compared with the critical value. If the calculated value of Chi-Square goodness of fit test was greater than the critical value, the null hypothesis was rejected and concluded that there is a significant difference between the observed and the expected frequency. If the calculated value of Chi-Square goodness of fit test was less than the table value, the null hypothesis was accepted and concluded that there is no significant difference between the observed and expected value:

- $\chi^2_{calc} < \chi^2_{krit}$ :  $H_0$  is accepted,
- $\chi^2_{calc} > \chi^2_{krit}$ :  $H_1$  is accepted

The degree of freedom is:

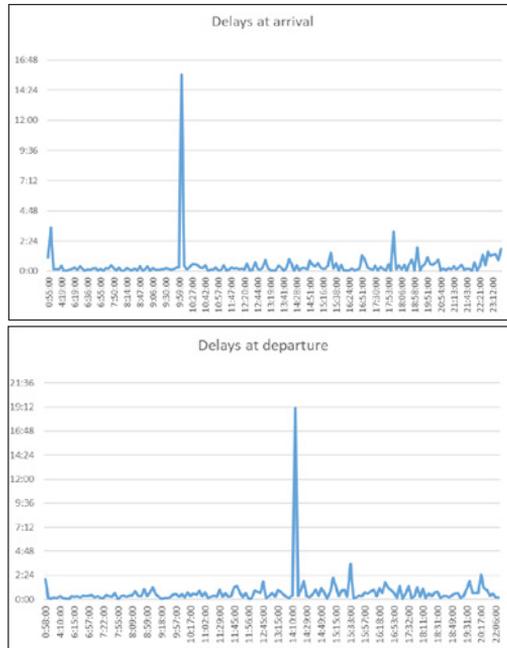
$$df = k - p - 1 \tag{9}$$

where:  $k$ : the number of classes with at least 5 frequency;  $p$ : is the number of estimated parameters for the distribution;  
The value of  $k$  is need to be decreased by the number of classes with less than 5 frequency. These classes were merged together.

### 3. Results

On the investigated day, there was 308 flights movement in total at the Budapest Liszt Ferenc International airport. From which almost equally 50%-50% was the distribution between the departing and arriving flight.

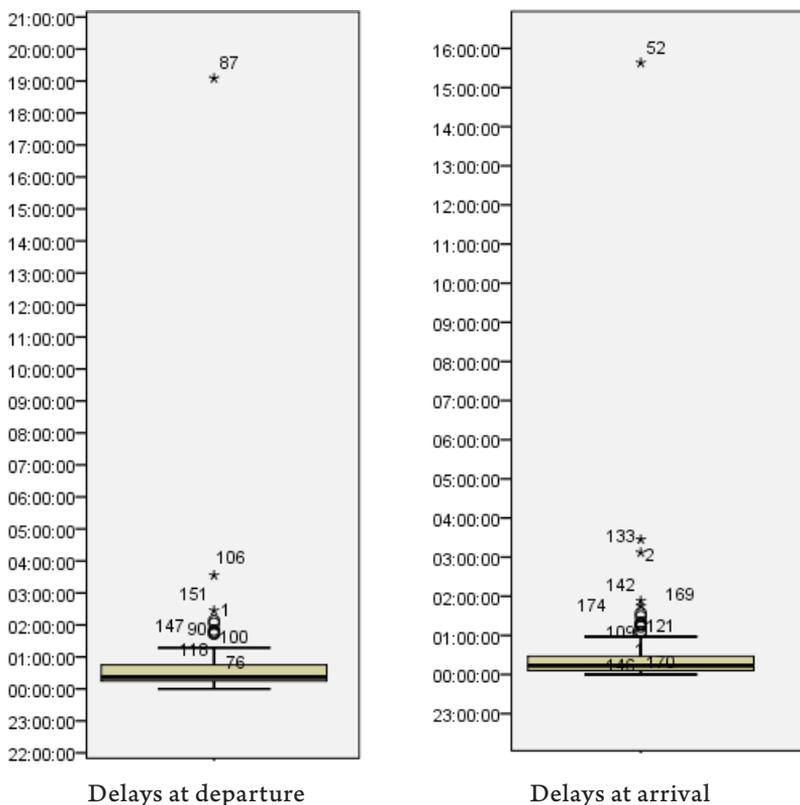
On Figure 2, randomly chosen workdays delay of departure and arrival is shown. The  $x$  axis is the scheduled departure or arrival time (from 00:00:00 to 24:00:00 calculated in minutes), and  $y$  axis is the delay in minutes.



**Fig. 2.**  
Delays Through the Day  
Source: own compilation based on dataset originated from HungaroControl Ltd.

An interesting phenomenon in the pattern could be investigated, that there are flights that are stand out apart from others. These flights may be originally scheduled to depart a day before, but the actual departure time is pushed to a day later, to

the considered day. Hence authors want to investigate regular daily distribution of delay therefore exclusion of extremities were necessary (see boxplot diagrams in Figure 3). On Figure 3. the extreme values are clearly identifiable.

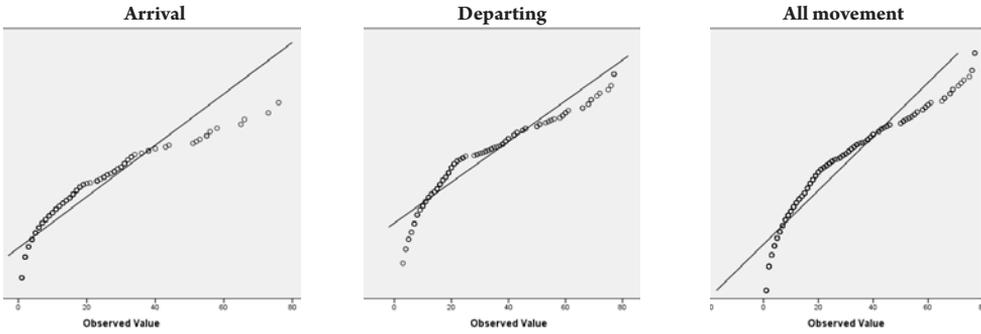


**Fig. 3.**  
 The Results of the Boxplot Graphs  
 Source: own research

### 3.1. Normality Tests

Figure 4 shows the normal probability plot of our data. From this we can conclude that the normality rule is violated. The

two data sets: the 45 degree angle and the Q-Q plot not fall on the reference line. This suggests that there is no normality between the plotted residuals and the expected values.



**Fig. 4.**  
*The Results of the Graphical Tests*  
 Source: own analysis

The results of the objective Chi-square goodness of fit test for Normality is presented in Table 2.

**Table 2**  
*Results of Chi-Square Test for Normality*

	Delay at Arrival	Delay at Departure	Delay of All Movements
<i>df</i>	19	19	16
$\chi^2_{calc}$	6110.00	6375.80	15314.65
$\chi^2_{krit}$ :	30.1435	30.1435	26.2962
Accepted hypothesis	H1	H1	H1

Source: own edition

#### 4. Conclusion

Based on the presented results, authors can clearly determine that the initial assumption that the deviation of the actual and the planned departing and arriving time at the airport follows Normal distribution is not confirmed. Neither the graphical nor the fitting test showed to any of the assessed statistical population to follow Normal distribution.

The showed method will allow the authors to automatize the process and be able to assess

longer period. Further on this analysis could be a solid basis of further infrastructure vulnerability research.

#### References

Čokorilo, O. 2008. Risk management implementation in aircraft accident cost analysis. In *Proceedings of the 12th Air Transport Research Society (ATRS) World Conference*, 98-102.

Čokorilo, O.; Dell'Acqua, G. 2013. Aviation Hazards Identification Using Safety Management System (SMS) Techniques. In *Proceedings of 16th International conference on transport science (ICTS 2013)*, 66-73.

- Čokorilo, O.; Gvozdenović, S.; Mirosavljević, P.,; Vasov, L. 2010. Impact of aircraft emissions on the environment, *Journal of Applied Engineering Science* 8(3): 123-138.
- Ford, C. 2018. Understanding Q-Q Plots. Available from Internet: <<https://data.library.virginia.edu/understanding-q-q-plots/>>.
- Ghasemi, A.; Zahediasl, S. 2012. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians, *International Journal of Endocrinology and Metabolism* 10(2): 486–489. doi: 10.5812/ijem.3505.
- Kling, F.; Somosi, V.; Pokorádi L.; Rohács D. 2017. Budapest Liszt Ferenc International Airport Aircraft Traffic Analysis with Markov Processes [in Hungarian: Budapest Liszt Ferenc Nemzetközi Repülőtér légi jármű forgalmának elemzése Markov-folyamatokkal], *Aviation Science Letters [in Hungarian: Repüléstudományi Közlemények]* 29(3): 115-126.
- Madácsi, R. 2015. Applying data science techniques to increase runway efficiency [in Hungarian: Data science technikák alkalmazása a futópálya-hatékonyság növelésében], *Aviation Science Letters [in Hungarian: Repüléstudományi Közlemények]* 27(3): 159-170.
- Mirosavljević, P.; Gvozdenović, S.; Čokorilo, O. 2011. A model of air traffic assignment as part of airport air pollution management system, *Aviation* 15(4): 92-100. doi: doi.org/10.3846/16487788.2011.651792.
- Sipos, T. 2017. Spatial Statistical Analysis of the Traffic Accidents, *Periodica Polytechnica Transportation Engineering* 45(2): 101-105. doi: 10.3311/PPtr.9895.
- Serhan, D.; Lee, H.; Yoon, S.W. 2018. Minimizing airline and passenger delay cost in airport surface and terminal airspace operations, *Journal of Air Transport Management* 73: 120-133. doi: 10.1016/j.jairtraman.2018.07.001.