CORRESPONDENCE ANALYSIS IN EXAMINATION OF REASONS FOR FLIGHT SCHEDULE PERTURBATIONS

MIRA PASKOTA*
OBRAD BABIĆ

University of Belgrade
The Faculty of Transport and Traffic Engineering

Abstract

This paper presents the results of analysis of reasons for flight schedule perturbations and the possibilities of their elimination. Correspondence analysis was used for statistical analysis on the JAT Airways database for reasons for departure delays and their length. There was indication of positive correlation between the reasons and the delays. But for more detailed insight into the relationship between departure delay length and their reasons, as well as the likelihood of their elimination, more homogenous data analysis was necessary (e.g. domestic flights, flights from Belgrade etc). After the proposed database analysis was completed it was concluded that there are some reasons for short departure delays (up to 45 minutes) which are the consequence of the JAT airline omissions. The number of these flight delays is considerable and they can produce serious flight schedule perturbations which could be eliminated with relatively ease by correcting these airline omissions.

Keywords: *Flight schedule perturbations*, *Flight delay*, *Correspondence Analysis*, *Multiple response tables*

* Corresponding author:
University of Belgrade
Faculty of Transport and Traffic Engineering
Vojvode Stepe 305
Belgrade, 11000 - Serbia
email: m.paskota@sf.bg.ac.yu
Tel: (+ 381 11) 3091330
Fax: (+ 381 11) 3096704

1. INTRODUCTION

Perturbations on a planned flight schedule arise for various reasons. These perturbations are usually manifested in the form of flight delays and/or flight cancellation. According to their source, the reasons for flight delays are usually divided into three groups: delays due to omissions by the airline (aircraft failure, passenger services, shortage of crew etc.), delays generated by other airport related agents (airports and air traffic control) and delays emerging as result of various environmental influences (meteorological conditions, military flights etc.). The reasons for delays can also be grouped by the delay source location: delays which occurred at the departure airport, delays arising during the flight or a mix of both. In many cases, delay of a particular flight is the consequence of several reasons that belong to different groups and/or sources.

Flight schedule perturbations are a serious problem for airlines, air traffic control, airports and other participants in the air transport industry. Numerous authors, among them Luo and Yu (1997) and Wu and Caves (2000, 2002), have examined this problem. Paskota et al (2005) analyzed flight schedule perturbation using departure delay length and flight type. In that paper only flight delays longer than 15 minutes were analyzed, because the paper deals with flight schedule perturbation from the point of view of existing flight schedule reliability and designing a new daily operational flight schedule in which the influence of perturbations on flight schedule reliability would be as minimal as possible.

However, even short departure delays of less than 15 minutes can significantly influence flight schedule performance. From the point of view of the dispatcher in an operational center responsible for minimizing flight schedule perturbations, even small perturbations can have a significant influence on flight schedule performance, especially taking into account the usually large number of those perturbations. Pakota et al (2005)'s analysis shows that there is a considerable number of small perturbation events, therefore, justifying the consideration of analysis of all flight departure delays, with special emphasis on the analysis of short delays.

The database on departure delay length and departure delay reasons, among other things, contains data that indicates that a flight delay is often the consequence of several reasons from the same or from different flight schedule perturbation events group and/or sources. From such a database, containing possibilities for multiple reasons, it is difficult to establish the connection between flight delay length and the individual group of reasons for flight delay using only standard statistical descriptive analysis and contingency tables analysis. Establishing such connection, if it exists, is of great importance, especially for the group of reasons of delays due to airline omissions, since the airline is in position to influence such perturbations directly, and eliminate them, entirely or in a large part.

The objective of this paper is to identify possible connections between departure delay length and the previously mentioned group and/or source of reasons for delays that are due to airline omissions. That is the reason why in this paper all flight departure delays experienced by JAT Airways during the year 2002 are analyzed, not just delays longer than 15 minutes, as was done by Paskota et al (2005). In order to verify and to describe the positive correlation between flight delay lengths and their source of reasons, Correspondence Analysis (Benzécri, 1973) was suggested and applied.

2. DATA DESCRIPTION

In this paper only flights which experienced departure delays and have information about the reason or reasons for delay are analyzed. This is due to the fact that data on the reasons for arrival delays is not available in the database, so it was impossible to analyze it in the same contexts as departure delays. Arrival delays of the aircraft at the destination airport could be caused by departure delays, delays generated during the flight or by a combination of both. There are examples of the lessening of the duration of delay that an aircraft initially had at the departure or, in some cases, even the complete compensation of that delay during the flight.

Unfortunately, in the analyzed database for the year 2002 (it contains 16,901 flights) even the data about the reasons for the departure delays is not complete, i.e. there are missing cases. Some flights that have experienced departure delays have no stated reason for that delay, so it was impossible to analyze them. The differences between the numbers of described sets of flights and a summary of the database is presented in Table 1. For example, during the year 2002 JAT Airways completed 7,929 flights from Belgrade, of which 1,521 experienced a departure delay. Of those 1,521 delayed flights from Belgrade, only 1,490 had information about the reason for their delays. However, there are 1,565 reported reasons for flight delays for these 1,490 flights from Belgrade (the flight could be delayed due to more than one reason, i.e. multiple responses are possible).

Table 1 Summary of database

Flights	Total number of flights in the database	Number of flights with departure delay	Number of flights with stated reason for departure delay	Total number of reasons for delay
All	16901	5220	4293	4893
European	11619	3926	3250	3769
Domestic	4250	874	661	701
From Paris	567	467	418	552
From Belgrade	7929	1521	1490	1565

The database of flight perturbation provides the departure delay length with the precision of minutes. For the purpose of the analysis in this paper, all delay lengths were divided into eight categories (Table 2).

Table 2 Categories of the variable *Delay duration*

Code	Delay duration
1	1 to 10 minutes
2	11 to 15 minutes
3	16 to 20 minutes
4	21 to 30 minutes
5	31 to 45 minutes
6	46 to 60 minutes
7	61 to 90 minutes
8	more than 90 minutes

In order to avoid confusion with the terminology, we first need to define the terms used throughout this paper:

- All delays – departure delays of aircraft greater than 0 minutes (all categories of the variable Delay duration)
- Delays – departure delays of aircraft greater than 15 minutes (categories 3 to 8 of the variable Delay duration)
- Short delays – departure delays of the aircraft between 1 and 45 minutes (categories 1 to 5 of the variable Delay duration)

In the original database reasons for delay are coded according to the AIC[2] code list into 66 different categories. In this paper these are grouped into 15 principal classes according to the same code list (Table 3).

Table 3 Categories of the variable *Reason for delay*

Code	Reason for delay
1	Miscellaneous
2	Passengers, baggage, cargo or mail
3	Aircraft servicing
4	Aircraft failure
5	Technical failure or equipment damage
6	Flight plan or crew
7	Meteorological conditions
8	ATFM due to ATC restrictions
9	ATFM other
10	Security
11	Airport services
12	Restrictions at the departure airport with or without ATFM restrictions
13	Short connection, flight crew rotation
14	Aircraft rotation
15	Unknown

In the case of domestic flights (flights where both departure and arrival airport are in Serbia and Montenegro) from 15 principal classes of delay reason only class 9 is omitted. In the

[2] AIC (Airline Internal Codes) code list is in use in the JAT Airways for coding of the reasons for the flight delays.

case of flights from Belgrade (departure airport is BEG - Belgrade) three classes are omitted: 1, 6 and 12. For flights from Paris to Belgrade (departure airport is airport CDG - Paris) seven classes are omitted: 1, 4, 5, 6, 7, 9 and 13.

Naturally, it is usual that a flight may be delayed due to more than one reason. In the database of the analyzed flights a maximum of three reasons for a flight delay are recorded. This is the reason why, during the analysis of this problem, we had to deal with the analysis of multiple responses tables.

3. CORRESPONDENCE ANALYSIS

Correspondence analysis (Benzécri, 1973; Hill, 1974; Greenacre, 1984; Paskota, 1999) is a well known method for multivariate analysis of nominal variables. Correspondence analysis is normally used for the analysis of the association between two nominal variables, which are two sets of binary variables. This is a very good methodology but unfortunately a rather neglected way of analyzing contingency tables. Usual way in categorical data analysis is to consider each nominal variable as one set of binary variables. De Leeuw (1973) proved that correspondence analysis could be regarded as canonical correlation analysis between two sets of binary variables and that singular values from correspondence analysis are coefficients of canonical correlation.

Canonical correlation is a well known term from multivariate statistics, and it has dual meaning. First meaning of term canonical correlation is to denote important method of multivariate statistical analysis suggested by Hotelling in his famous paper (Hotelling, 1936). This method allows the analysis of association between two sets of numerical variables. Second meaning of term canonical correlation is to describe coefficients that measure intensity of association between two sets of variables as a final result of application of canonical correlation as a method. One of the results of canonical correlation analysis is the set of coefficients that measure the intensity of association between latent dimensions from two sets of variables. Because of similarity to product moment correlation coefficient, and being the result of canonical analysis of the matrix, those coefficients were named canonical correlations.

Latent dimensions obtained as result of correspondence analysis are common for the space determined by rows of a contingency table, as well as for the space determined by columns of a contingency table. Maximal number of dimensions of this space is equal to the rank of analyzed matrix, i.e. for one is less than the minimum between the number of categories of variable used to determine the rows and number of categories of variable used to determine columns of the analyzed table. Latent dimensions are not of equal significance, so in practice one who analyzes the data usually discards some of them and keeps fewer dimensions in the further analysis (most often 2 or 3). In that manner, if not absolutely accurately, the positions of categories in latent space are well enough represented.

One of steps in correspondence analysis is the calculation of coordinates according to all latent dimensions for all categories of both variables. The categories of the variable that represent rows of the table could be graphically presented in the latent space of wanted number of dimensions. In the same way, the categories of the variable that represent columns of the table could be graphically presented in the latent space of wanted number of dimensions. Because that space is common for both sets of categories, an usual way of presenting the results of correspondence analysis is to use so cold biplot, i.e. graphical representation of positions of both variable categories in common latent space of small dimension (usually 2 or 3).

When interpreting results, distances between categories and coordinate axes are used to represent latent dimensions, as well as distances between the categories of the same variable. It is not possible the interpretation of the distances between the categories of different variables, regardless of the fact that those points are graphically presented in the same space. The relationship between the categories of two variables might be determined indirectly using positions to the common coordinate axes and the meaning of those axes.

Of the two analyzed variables, *Reason for delay* is an entirely nominal variable, while *Delay length* is ordinal, i.e. an ordered categorical variable. In the context of this analysis both variables are regarded as and treated in the same level – as nominal variables. Since, as already stated, multiple responses are possible to explain delay reasons, the resulting tables are not ordinary contingency tables, but multiple response tables. Contrary to the other methods common in the analysis of contingency tables, which are not suitable for cases like this, correspondence analysis gives excellent results in the analysis of multiple responses tables (Nishisato, 1993).

The variable *Delay length* consists of 8 categories, while the number of categories of variable *Reason for delay* depends on the analyzed set of data (i.e. type of flights) and ranges between 8 and 15. The dimensions of analyzed tables are: for all flights 8×15, for European flights 8×15, for domestic flights 8×14, for flights from Paris to Belgrade 8×8 and for flights from Belgrade 8×12. Because of the smaller dimension in all tables is equal to 8 (number of categories of the variable *Delay length*), 7 is the maximum number of dimensions in the solution for all cases of analysis.

Table 4 Canonical correlations and inertia for all, European, domestic and flights from Paris to Belgrade

	All flights			**European flights**			**Domestic flights**			Flights from Paris to Belgrade		
Dim.	**cc**	**Inertia**		**Cc**	**Inertia**		**Cc**	**Inertia**		cc	Inertia	
1	.333	.722	.722	.329	.719	.719	.457	.546	.546	.263	.563	.563
2	.153	.152	.874	.137	.124	.844	.306	.244	.790	.157	.200	.763
3	.081	.042	.916	.100	.067	.911	.177	.082	.872	.130	.138	.901
4	.076	.038	.954	.083	.046	.956	.150	.058	.930	.083	.056	.957
5	.068	.030	.984	.069	.032	.988	.128	.043	.973	.064	.033	.991
6	.039	.010	.994	.033	.007	.996	.089	.021	.994	.029	.007	.997
7	.031	.006	1.00	.026	.004	1.00	.048	.006	1.00	.018	.003	1.00

From Table 4 we can see that the association between the variables *Delay length* and *Reason for delay* is the strongest in case of domestic flights (first canonical correlation of 0.457, second of 0.306 and total of five canonical correlations greater than 0.1). This association is the weakest in case of the flights from Paris (first canonical correlation of 0.263). In case of flights from Belgrade (Table 5) this association is also very strong; the first canonical correlation is equal to 0.420. From the values of the inertias from the Tables 4 and 5, it follows that a two-dimensional graphical representation of results explains 87.4% of total variability in the case of all flights, and 89.1% of total variability in the case of flights from Belgrade.

In case of short delays from Belgrade (Table 5) the number of categories of the variable *Delay duration* is 5, so that the maximum number of dimensions in the solution is equal to 4. In this analysis the first canonical correlation is relatively low (0.236) and two-dimensional graphical representation of results in this case explains 84.7% of total variability.

Table 5 Canonical correlations and inertias for the flights from Belgrade

	All delays			**Short delays (≤ 45 min)**		
Dim.	**cc**	**Inertia**		**Cc**	**Inertia**	
1	.420	.771	.771	.236	.751	.751
2	.165	.120	.891	.085	.097	.847
3	.104	.047	.938	.082	.091	.938
4	.084	.031	.969	.068	.062	1.00
5	.066	.019	.988			
6	.046	.009	.997			
7	.027	.003	1.00			

4. ANALYSIS OF ALL FLIGHTS AND ALL THEIR DELAYS

Two-dimensional graphical representation of the correspondence analysis results for all flights is presented in Figure 1.

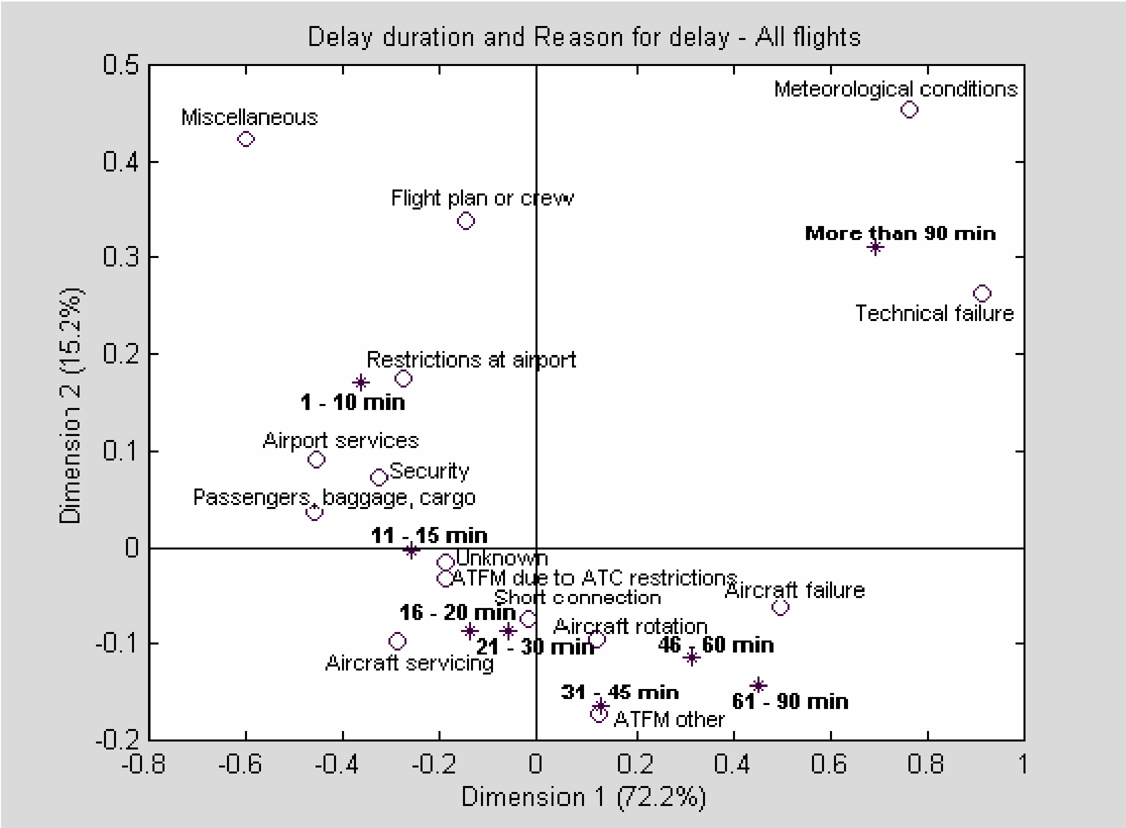

Figure 1 Two-dimensional graphical representation of correspondence analysis results for all flights

The interpretation of the results presented in Figure 1 is as follows: The first dimension explains 72.2% of the total variability of the data from multiple response tables. With no doubt, dimension could be interpreted as the time axis with shorter delays positioned on the left side and gradually incising towards the longer delays positioned on the right side of a two-dimensional graphical representation. The second dimension is not as easy to interpret and in order to do so it is necessary to further explore the meaning of particular categories of the variable *Reason for delay*. As can be seen in Figure 1, in the upper part of the graph (positive values of the second dimension) besides the reason for the delay category *Miscellaneous* are also the following categories: *Meteorological conditions*, *Flight plan or crew*, *Technical failure or damage of the equipment* and *Airport services*. In the lower part of the graph (negative values of the second dimension) are categories: *ATFM other*, *Aircraft rotation*, *Short connection*, *Aircraft servicing* and *Aircraft failure*. The contrast between these two sets, containing reasons from different groups for aircraft delays, does not lead to an obvious and logical explanation. The reason why interpretation of the second dimension is difficult lies in the fact that, in the case of all flights, we are dealing with very heterogeneous data, thus providing a mixed set of reasons for the emergence of delays, obviously different for different classes of flights. In order to eliminate this problem, more homogenous sets of data delays of domestic flights and delays of the flights from Belgrade were analyzed.

5. ANALYSIS OF ALL DOMESTIC FLIGHT DELAYS

The two-dimensional graphical representation of correspondence analysis results for domestic flights is presented in Figure 2.

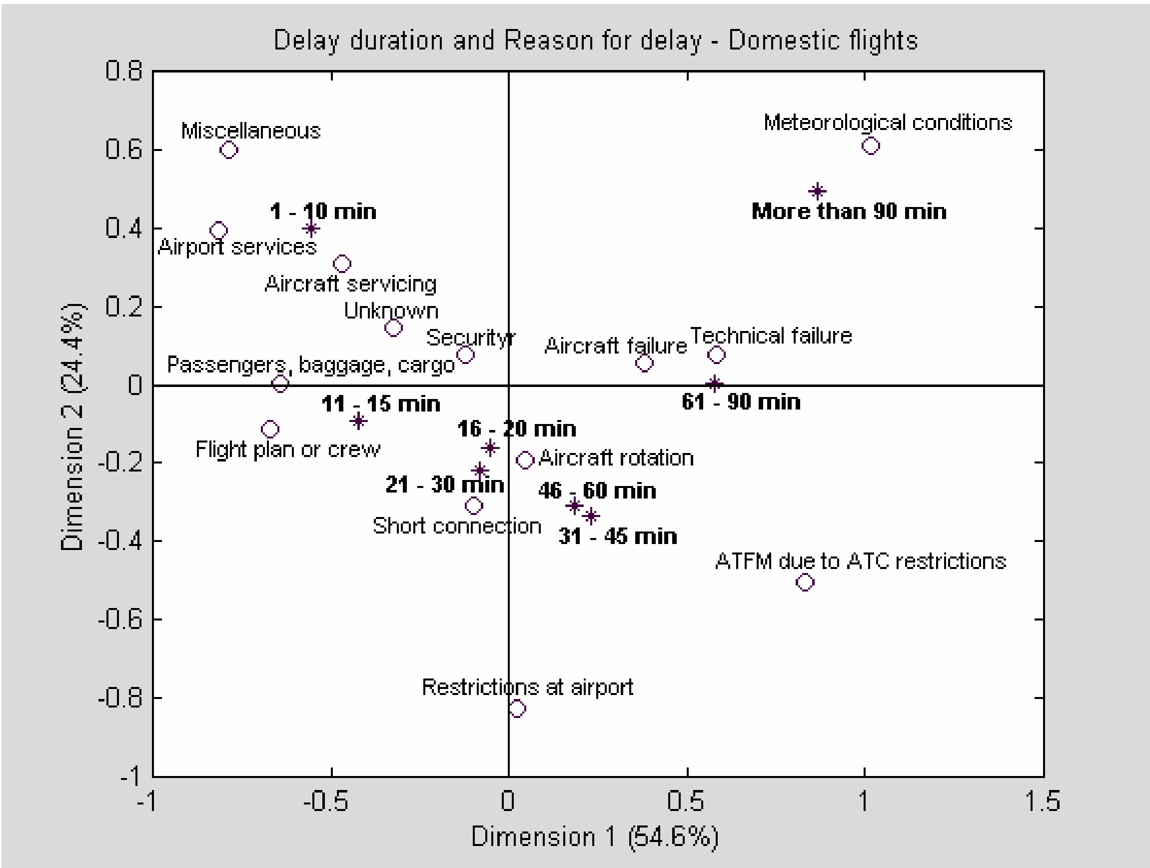

Figure 2 Two-dimensional graphical representation of correspondence analysis results for domestic flights

Interpretation of the results presented in Figure 2 is as follows: *Reason for delay wise,* domestic flights are more homogeneous than all flights or European flights, which is shown by the greater first canonical correlation (Table 4). This result is in accordance with the expectation due to the fact that domestic flights are defined as flights with both the departure and arrival airports within the territory of Serbia and Montenegro.

In case of the analysis of domestic flights, the first dimension explains 54.6% of the total variability and, as was the case with the all flights, is easily recognized as the time axis. Since the data on domestic flights is more homogenous than that on all flights, in this case the second dimension is somewhat easier to explain. As can be seen from Figure 2, on the upper parts of the graphical representation (positive values of the second dimension), as well as the category *Miscellaneous* are also the following categories: *Meteorological conditions*, *Airport services*, *Aircraft servicing*, *Security*, *Aircraft failure*, *Technical failure or equipment damage* and *Passengers, baggage, cargo or mail*. On the lower parts of the graphical representation (negative values of second dimension) are categories *Restrictions at airport of departure with or without ATFM restrictions*, *ATFM due to ATC restrictions*, *Short connection* and *Aircraft rotation*. It is easy to observe that the reasons for delays from the first group are of local character, i.e. common for all of them is that they emerge right at the airport of departure, while the reasons for delays from the second group in fact originate on some other place and only manifest themselves at the airport of departure. Consequently, it is logical to interpret the second dimension as the axis that explains the place of origin of flight delay.

6. ANALYSIS OF FLIGHT DELAYS FROM BELGRADE

Flights from Belgrade are defined as all flights with departure from Belgrade Airport, regardless of the location of the arrival airport: within the territory of Serbia and Montenegro, in European countries or outside Europe. In the initial stages of the analysis it was uncertain if domestic flights or flights from Belgrade are more homogeneous regarding to the reason for delay.

The two-dimensional graphical representation of correspondence analysis results for all delays from Belgrade is presented in Figure 3.

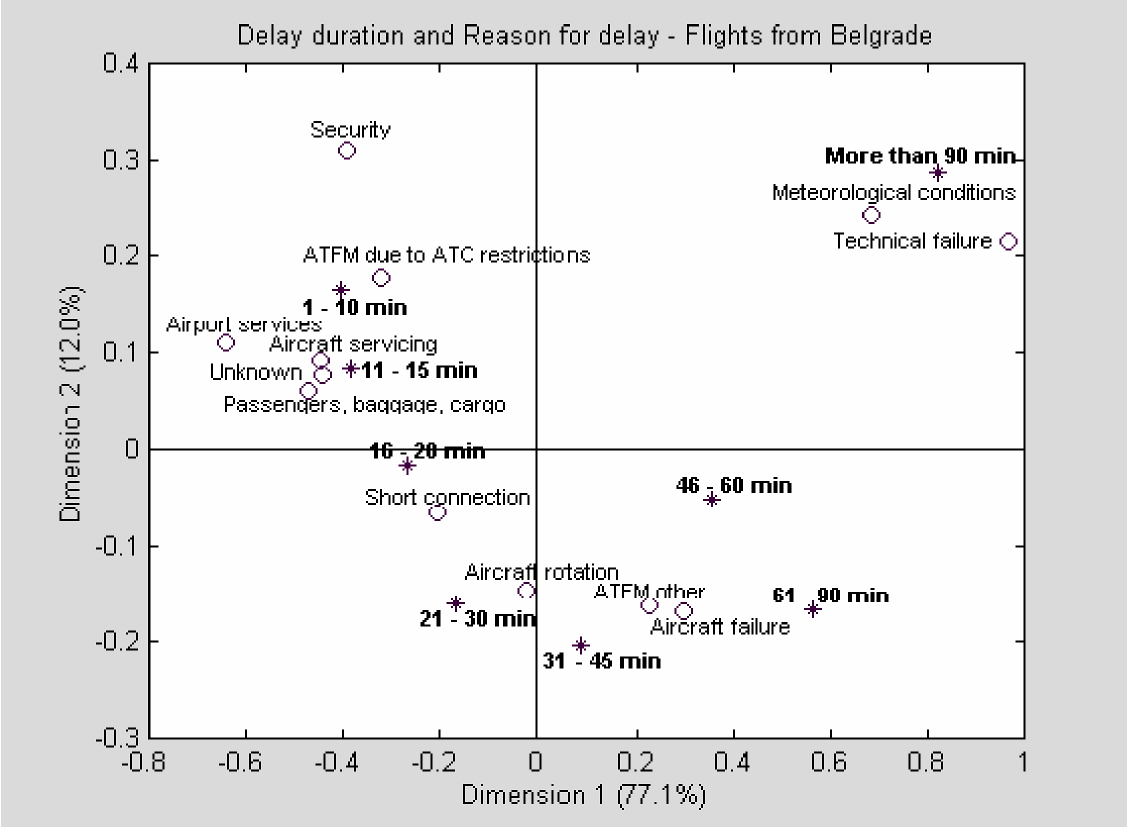

Figure 3 Two-dimensional graphical representation of correspondence analysis results for all delays from Belgrade

At first glance Figure 3 supports the thesis that flights from Belgrade are more homogeneous than domestic flights. This is an interesting conclusion, bearing in mind that the number of flights from Belgrade is two times larger than the number of domestic flights and that arrival airports are numerous and distributed all over the world. This is not the case with the other flights from airports within the territory of Serbia and Montenegro, because flights from those airports have a relatively small number of arrival airports.

In case of the analysis of flights from Belgrade, the first dimension explains 77.1% of the total variability and is easily recognized as the time axis. As was the case with the previous analysis, in this case too the interpretation of the second dimension is more interesting, which explains 12% of the variability. As can be seen in Table 5, of all correspondence analyses presented in this paper, the total variability explained by the two-dimensional space is the greatest in the analysis of flights from Belgrade (89.1%).

In Figure 3 is easy to notice that the categories of the variable *Reason for delay* in position follow the categories of the variable *Delay length* and are grouped around them. The first two categories of the variable *Delay length* (1 to 10 minutes and 11 to 15 minutes) are positioned very closely in the graph, so it is possible to conclude that in the light of this analysis there is no important difference between them. Near to these categories are placed the reasons that cause the shortest delays: *ATFM due to ATC restrictions*, *Airport services*, *Aircraft servicing* and *Passengers, baggage, cargo or mail*. Somewhat longer delays, from 16 to 20 minutes, are caused by *Short connection* and *Flight plan or crew*. Reasons for delays from 21 to 60 minutes are *Aircraft rotation*, *ATFM other* and *Aircraft failure*. The longest delays, ones greater than 90 minutes are due to *Technical failure, damage of the equipment* or *Meteorological conditions*.

From the previous analysis it can be concluded that short departure delays (less than 45 minutes) are generated on and are connected to the departure airport, in this case, the Belgrade Airport. These delays influence significantly the perturbations of flight schedules and the reasons for these delays could somewhat be eliminated. So it is very important to investigate the reasons for short delays of aircraft departing from Belgrade Airport.

On this occasion only short departure (delays up to 45 minutes) are analyzed in detail (Figure 4).

From Figure 4 it is clear that reasons for delay are uniformly distributed in the space determined by the first two dimensions. The reasons that cause the longest delays (*Meteorological conditions*, *Technical failure or damage of the equipment* and *Aircraft defects*) are positioned at the right side of the two-dimensional graphical representation and do not deviate significantly relatively to the second dimension. The reasons for the shortest delays (*Airport services* and *ATFM due to ATC restrictions*) are positioned at the left side of the two-dimensional graphical representation and also do not deviate significantly relatively to the second dimension. However, it is possible to conclude that the shortest and probably negligible delays (shorter than 10 minutes) emerged because of the following reasons: *Passengers, baggage, cargo or mail* and *Aircraft servicing*, while *Security* causes slightly longer delays. Medium delays, from 16 to 20 minutes, are caused by *Short connection* and *Aircraft rotation*, obviously both reasons originating with the airline (in this case JAT Airways). Those delays could easily be eliminated by an airline, for example, by applying a longer flight block time and/or longer ground handling time for certain flights. The delays between 21 and 45 minutes are generated by the ATFM and those delays cannot be eliminated easily by an airline.

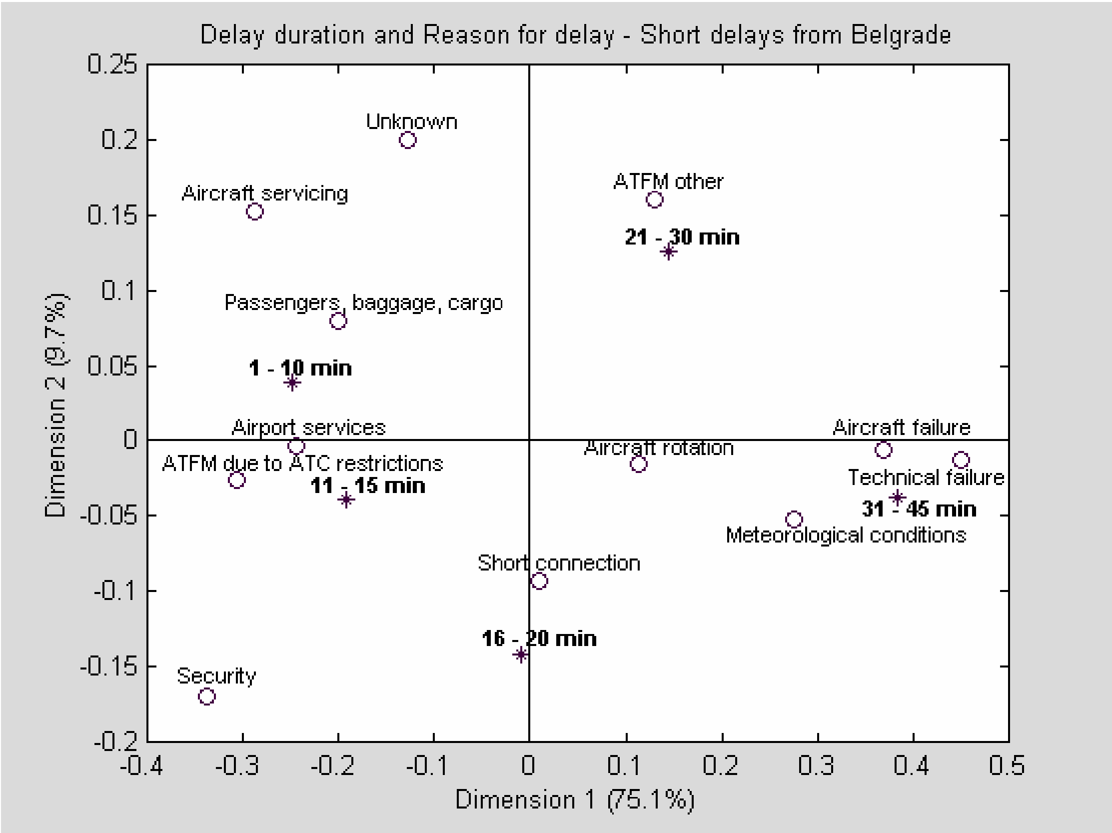

Figure 4 Two-dimensional graphical representation of correspondence analysis results for short delays from Belgrade

7. CONCLUSIONS

Detailed analysis of the reasons for departure flight delays and the length of those delays for JAT Airways, according to the correspondence analysis method presented in this paper, lead to numerous conclusions. First of all, it is not possible to talk about a relation between source and length of a flight delay in general. Some general rules do certainly exist, as is shown by the results of the analysis of all flights (Table 4 and Figure 1). But these are well known and expected facts that do not offer any new information to a manager responsible for dealing with flight schedule perturbations, regardless of the segment in air traffic industry that he/she represents: airline, airport or Air Traffic Control. This well-known conclusion is the consequence of the heterogeneity of the types of flights with departure delays (domestic, international flights etc.).

It would be desirable if a general rule existed and could be applied in all cases of flight delays. Unfortunately that is not the case. In order to acquire useful information for flight schedule designing improvement that will eliminate as many flight schedule perturbations as possible, it is necessary to concentrate on certain types of flights and analyze each of them independently, as the case of this paper.

The analysis of domestic flights further improved the understanding of the problem, pointing out the fact that flight delay problems are grouped according to the place of origin of the problem (airport of departure, ATFM, meteorological conditions, previous flights in the rotation). This enables one to solve the easier problem first i.e. to eliminate the delays of local character, those emerging at the departure airport (*Aircraft servicing*, *Passengers, baggage, cargo or mail*, *Airport services* and *Flight plan or crew*). It is not realistic to admit that single actors (airline, airport etc.) are able to solve the problem of delays that are originated in other places. So it is obvious that they have to combine their strengths and efforts at the national level, and also at the international level, in order to contribute to the solution of this problem at the global level.

The analysis of flights departing from the Belgrade Airport further clarified previously noticed patterns. The conclusion is that accidental delays (because of the various services - passengers, baggage etc.) are usually shorter than 10 minutes and are not something to worry about. In most cases those delays can be minimized either during the flight, or by reducing ground handling time. Long departure delays, of more than 90 minutes, of course are a serious problem, but those delays in this analysis occur relatively rarely and their solution must be provided by the dispatchers.

The major concern is the great number of medium length delays, which occur due to short connection time, flight crew substitution and aircraft rotation. Those delays have a significant influence on flight schedule perturbations and are usually caused by bad planning and design omissions by the airline. These are the problems that should be solved, not in an ad-hoc way, but through a systematic, well planned strategy and by consistent implementation by the airline.

The results of the correspondence analysis of several different sets of data presented in this paper show that the relationship between certain reasons and departure delay length does exist and that that relationship should be observed during the flight schedule designing process. One possible solution could be the appliance of longer flight block time or longer ground handling time for a certain number of flights. With this measurement most of the short delays would be eliminated. This is very important in a "hub and spoke" network where short delays can significantly influence flight schedule perturbations.

REFERENCES

Benzécri, J.P. (1973). *L'Analyse des Données, Tome 2: Correspondances*, Dunod, Paris.

De Leeuw, J. (1973). *Canonical Analysis of Categorical Data*, Leiden: University of Leiden (dissertation).

Hill, M.O. (1974). "Correspondence Analysis: A Neglected Multivariate Method", *Applied Statistics* 23, 340-354.

Hotelling, H. (1936). "Relations between two sets of variables", *Biometrika* 28, 321-377.

Greenacre, M. J. (1984). Theory and Applications of Correspondence Analysis, Academic Press, London.

Luo, S., Yu, G. (1997). On the airline schedule perturbation problem caused by the ground delay program, *Transportation Science* 31 (4), 298-311.

Nishisato, S. (1993). "On Quantifying Different Types of Categorical Data", *Psychometrika* 58, 617-629.

Paskota, M. (1999). *Multivarijaciona statistička analiza kategorijskih podataka*, Magistarska teza, Fakultet organizacionih nauka, Beograd.

Paskota, M., Babić, O., Kalić, M. (2005). "Analiza poremećaja reda letenja na osnovu baze realizovanih letova JAT-a u 2002. godini", *Tehnika LX*, 11-26.

Wu, C.L., Caves, R.E. (2000). Aircraft operational costs and turn-around efficiency at airports, *Journal of Air Transport Management* 6, 201-208.

Wu, C.L., Caves, R.E. (2002). Towards the optimization of the schedule reliability of aircraft rotations, *Journal of Air Transport Management* 8, 419-426.