A Predictive Model to Evaluate and Improve Punctuality of Fleet of Wide-body Aircraft Maintenance and Reliability

Vera Afonso Alves

Abstract

This research was carried with the aim of analyzing Martinair M&E impact on the network performance and providing an efficient tool to predict delays and the On-Time Performance (OTP).

The methodology undertaken in this research is based on conventional analytical methods to identify the 'real' delay root causes and their relative importance. The growth of delays and unplanned ground times are analysed in different ways, not only using rates and duration but also per month, per quarter, per year, per station and per Maintenance Delay Categorization Groups (MDCG). Fishbone diagrams are drawn to completely identify and understand the root causes. Statistical analysis (Binomial Logistic Regression and GLM-ANOVA) is applied to investigate the significance of the maintenance factors and their interactions and to build a regression equation which allows the development of the predictive model. The predictive model gives valid OTP results and enables managers and engineers to take preventive measures in order to enhance punctuality.

Keywords: Aircraft Delay, Punctuality, On Time Performance, Dispatch Reliability, Aircraft Maintenance

1. Introduction

In airline business there is an urge to measure the effectiveness of the maintenance process in order to find the best trade-off between safety and costs. According to Niehues et al [1], delay costs can be from 0.6% to up as much as 2.9% of airlines revenues and one way to increase the operating profit is to pay more attention to On-Time Performance (OTP).

For this purpose, the M&E main objective is to maximize the operational reliability and aircraft availability. Therefore, Martinair has already a *Reliability Program* with the aim of reducing the number of chargeable delays and unscheduled ground times (UGTs). However, the analysis of the chargeable delays is not sufficient if Martinair wants to be in control of its punctuality. It is essential to find out the bottlenecks of the processes. In other words, the next step for the Reliability Engineering is to analyze the impact of M&E on the network performance and implement measures to enhance OTP.

Hence, the objective of this research can be formulated as follows:

Develop a predictive model to perform analysis of Martinair M&E processes effectiveness with the aim of controlling/improving punctuality.

Other questions that cope with this objective need to be addressed:

1. What are the drivers for the delays and UGTs?
2. How can delays and UGTs be reduced?

2. Literature Review

2.1 Flight Delay and Dispatch Reliability

A complex chain of events occurs before aircraft departure and some of them may cause an unexpected delay. Sometimes a delay results from a single reason, but most delays come from multiple causes. The departure delay has increased significantly in the past decade due to several factors such as the increasing demand of air transport [2]. The International Air Transport Association (IATA) created the *IATA Delay Codes* to help airlines standardize the reason of a flight late departure. Furthermore, it is possible to group the delay factors under controllable and uncontrollable and also by airline activity as presented in the chart below [4].

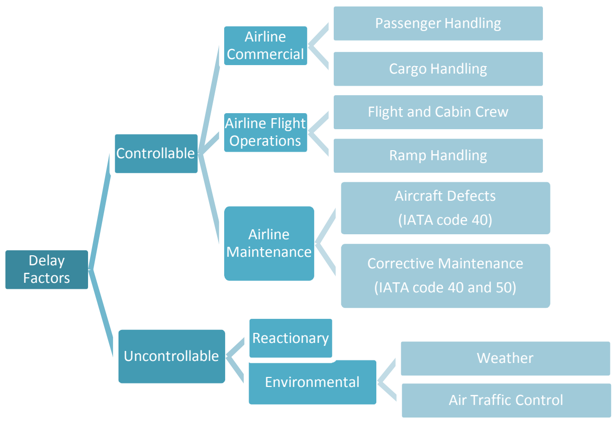

Figure 1 - Delay Factors based on the Boeing Seminar "Delay: How to Mitigate it"

The Maintenance & Engineering (M&E) department of an airline is directly responsible for the IATA delay codes 40 (Technical and Aircraft Equipment) and 50 (Damage to Aircraft and Automated Equipment Failure/EDP) which can be grouped as controllable factors. Therefore, it is important for M&E to find the root causes of the *Airline Maintenance* delay factors. This helps airlines to understand how they can take effective measures to be in control of their On-Time Performance. OTP, also known as, *Dispatch Reliability* is defined as the percentage of revenue departures which do not incur a delay because of technical problems [3]

$$Dispatch\ Reliability = OTP = \left(1 - \frac{\#\ of\ delays}{\#\ departures}\right) x\ 100 \quad (1)$$

Sridhar [4] mentioned that there are also influencing factors that can affect the *Dispatch Reliability.* The ones that are related to M&E and are interesting for further analysis are: *Age, Flight Length, Utilization* and *Airline*.

Age

Ageing aircraft can be described as a process through which the aircraft experiences the effects of an increasing age [5] It is further complicated to determine the age of an aircraft by the fact that individual aircraft components age differently depending on those factors. It was chosen in the current study not to take into account the age as a factor of On-Time Performance (OTP). A careful examination of the relation between aircraft age and dispatch reliability could be the subject for another research.

Flight Length

In aviation business, flight length is defined as the airborne time during a flight. This factor is not taking into account in the current research because Martinair, operates mainly cargo and passenger long-haul flights.

Utilization

Aircraft utilization is the average daily airborne flying hours or cycles for one aircraft. The aircraft utilization is calculated by the following equation:

$$Aircraf\ Utilization\ Rate\ (AUR) = \frac{Totalflying\ hours}{Number\ of\ days} \quad (2)$$

AUR represents the average daily utilization of an aircraft for the past 7 days.

Airline

The controllable factors can be related with the different airline departments: *Commercial, Flight Operations* and *Maintenance.* The basic structure of Martinair Maintenance & Engineering (M&E) is formed by four main departments: *Maintenance Control, Aircraft Maintenance & Supporting Shops, Engineering* and *Material Unit*.

2.2 Delay Predictive Models

In the literature, it is possible to find some studies that develop prediction model for the delays. Most of them are related to: prediction of delay propagation in the flight scheduled [6], [7], [8], aircraft rotations between airports [9], [10], [11], or prediction of delays from delay statistics of airports [12], [13]. The focus of these models is in airline and airport operation rather than in maintenance.

Mueller and Chatterji [13] used 12 variables to describe the distribution of delay. The Normal and Poisson distribution were used to compare with historical data and the Least Square method was also applied to good-fit the parameters. Yufeng et al [12] also developed a model for estimating the departure delay distributions using nonparametric methods for daily and seasonal trends. They use flight data from United Airlines and Denver International Airport to build a strategic departure delay prediction model.

Wu and Caves [11] explored the inherent delays of airline schedules resulting from limited buffer times and stochastic disruptions in airline operations. The results show that airline schedules must consider the stochastic nature of daily operations. One way of improve schedules reliability is to embed and design buffer times in airline schedules. Also Wu and Caves [10] investigated the relationship between tight schedule punctuality and aircraft turnaround efficiency at airports. Again, they developed a cost minimization model to optimize the scheduling of aircraft rotation by balancing the use of schedule time and delay costs [9].

Another approach by Sachon and Pate [14] developed a model to assess the effects of an airline maintenance policy on delays, cancellation and in-flight safety. They propose a probabilistic risk analysis model which consists of three tiers: management policies ad decisions (Decision tier), maintenance and delays (Ground model tier) and in-flight safety (In-Flight model tier).

Abdelghany et al [8] tried to use the classic shortest path algorithm to model and predict the flight scheduled delay for United Airlines of U.S. The model used a directed acyclic graph containing a series of nodes, which were sorted topologically in a liner time, to represent and simulate the process of scheduled flights operation and delay propagation. This model has been used at United Airlines' Operation Control Centre to monitor the daily schedule operation. AhmadBeygi et al [6] and [7] showed how delay propagation can be reduced by redistributing existing slack in the planning process, making minor modifications to the flight schedule while leaving the original fleeting and crew scheduling decisions unchanged. They used data from a major U.S. carrier to present computational results.

2.3 Statistical Analysis

Statistical analysis is based on the proposition that the behaviour of a dependent variable can be explained by a model that takes a form of an algebraic equation that involves other independent variables that describe the experimental conditions [18]. For a data analysis, after obtaining the parameters estimates, it is very important to understand how well the model actually fits the observed data.

Variables

There are dependent variables called response variable and independent variables called as factors or predictors. These variables can be from two types: *categorical variables* or *continuous variables.* The first ones, *categorical variables*, are variables that can only assume a limit number of possible values known as factors levels. *Continuous variables* are, as the name suggests, variables of a continuous function that can assume any value between a lower and upper limit [19].

Regression Analysis

In its simplest form regression analysis involves finding the best straight line relationship to explain how the variation in the dependent variable depends on the variation in the independent variable. Thus, regression analysis is the technique most used to analyze quantitative data and make forecast [20].

In general, the multiple regression procedures will estimate a linear equation of the form [22]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon \quad (3)$$

where

y dependent variable

x_i independent variables with i=1,..., m

m number of independent variables

β_0 intercept (value when all the independent variables are zero)

β_i i^{th} regression coefficient with i=1,..., m

ε error of prediction with mean zero and variance σ^2

Binary Logistic Regression

Logistic Regression analysis extends the techniques of multiple regression analysis to research situations in which the outcome variable is categorical like a binary response variable. It is based on probabilities associated with the values of the response [24]. In theory, the hypothetical, population of cases for which Y=1 is defined as p=P (Y=1). Then the theoretical proportion of cases for which Y=0 is 1-p=P(Y=0). In regression context, it is assumed that there is a set of predictor variables X_1,..., X_m (m independent variables) that are related to Y and, therefore, provide additional information for predicting Y.

In statistics, logistic regression is used to predict the probability of occurrence of an event by fitting data to a logistic curve. Then, this type of analysis is based on a linear model for the natural logarithm of the odds. Odds are more familiar to most people and represent the relative frequency with which the outcome occurs. Odds are directly related to probabilities and can be translated back and forth using the following relation [19].

$$Odds\ ratio = \frac{p}{1-p} \quad (4)$$

where p is a conditional probability of the form P(Y=1| X_1,..., X_m).

The logit transformation used in this analysis is done by taking the natural log of the odds [22]. A logit transformations is *link function* which means that it maps the interval (0, 1) onto the entire real number [23]. Then, the Logistic Regression analysis model is identical to the multiple regression analysis one, as shown in equation 2.10.

$$Log_e\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m \quad (5)$$

This analysis will find out the coefficients for the equation 5. After that it is possible to convert from a log-odds ratio to a probability by using the equation 6

$$p = P(Y = 1|X_1, \ldots, X_m) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m}} \quad (6)$$

Model Validation

For the GLM-ANOVA analysis, which will use the *delay time* as a dependent variable, we must check the value of the R^2 and *Adjusted* R^2. According to Miles and Shevlin [19], the R^2 represents the total amount of variance accounted for in the dependent variable by the independent variables. It is expressed as a percentage of the response variable variation that is explained by its relationship with one or more predictor variables.

The Binary Logistic regression analysis estimates the regression coefficients based on the *maximum likelihood principle*. As the name suggests, the estimator will be the value of the parameter that maximizes the likelihood function, L. This function is the product of the probabilities for delay (coded as 1) or for on time (coded as 0). The statistical software compute the regression coefficients which allows the calculation of the likelihood value. If it is considered the logarithm, the likelihood function can be written as follows.

$$log - likelihood = \sum_{i=1}^{n}\left[Y_i \log(p_i) + (1 + +Y_i) \log(1 - p_i)\right] \quad (7)$$

where $p_i = \left(\frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m}}\right)$ and Y_i is the real value of the response.

The objective is to discover the model that maximizes this value. As this value is always negative, we are looking for the closest number to zero.

3. Descriptive Statistics

3.1 Methodology

The delay data was gathered from January 2006 until April 2009 for B767 and MD11 fleet. When creating a database, it is important to choose the relevant information for the desire objective. It was used several data elements as Sridhar suggests [16]. The element *Delay Reason*, uses new categories called Maintenance Delay Categorization Groups (MDCG) which are a combination between the delay reasons used in Martinair and the ones presented by Niehues et al [1] and Shridhar [4]. The *Delay Reason* is divided in eight main groups:

- ✓ A/C Mechanical Fault
- ✓ Maintenance
- ✓ Ground Activities
- ✓ Non-Technical
- ✓ Down Line
- ✓ Parts
- ✓ FOD
- ✓ Incomplete Record

Furthermore, each of these 1st Order reasons has several specific reasons (2nd Order reasons).

Before continue, let's make clear the difference between *delay* and *unscheduled ground times (UGTs)*. As defined above, a *delay* occurs when a flight departure is later than the scheduled departure time due to a technical problem. An *UGT* happens when the period of time for maintenance inspections was not scheduled. In reality, the flight crew should send a list of all defects/alarms noticed during the flight an hour before arrival to the station of arrival and to the Maintenance Control (MC). It can be related to the extension of planned maintenance and aircraft damage.

Several graphs are plotted with the aim of getting an overall picture of the growth of the flight delays and UGTs of Martinair. The plotting analysis differentiates between delays and UGTs' graphs and, instead of using total numbers it takes into account rates and average time of delays. Thus, it is possible to compare fairly the on-time performance for different months/quarters/years. Otherwise it could lead to wrong interpretation of the data. The definitions of *Delay Rate (DR)*, *Average Delay Time (ADT)*, *UGT Rate (UR)* and *Average UGT (AUT)* are as follows:

$$Delay\ Rate\ (DR) = \frac{number\ of\ delays\ for\ a\ specific\ period}{number\ of\ departures\ for\ the\ same\ period}\ x\ 100 \quad (8)$$

$$Average\ Delay\ Time\ (ADT) = \frac{total\ time\ of\ delay\ for\ a\ specific\ period}{number\ of\ delays\ for\ the\ same\ period} \quad (9)$$

$$UGT\ Rate\ (UR) = \frac{number\ of\ UGTs\ for\ a\ specific\ period}{number\ of\ departures\ for\ the\ same\ period}\ x\ 100 \quad (10)$$

$$Average\ UGT\ (AUT) = \frac{total\ time\ of\ UGT\ for\ a\ specific\ period}{number\ of\ UGT\ for\ the\ same\ period} \quad (11)$$

This analysis looks not only for the departure in Amsterdam Airport Schiphol, which is Martinair's major station, but also for the worldwide business (including outstations). Moreover, it tries to understand which the most significant MDCG reasons are. As there are several reasons it must be applied a criteria to decide the most relevant that should be studied.

1. Delay Rate > 5% OR
2. UGT Rate > 5%

3.2 Results

For the sake of space, the graphs are not shown in this paper and it is only given the most relevant conclusions.

Worldwide and Schiphol

Martinair operates both passenger and cargo flights to several destinations around the world. Its base station is in Schiphol as well as its headquarters.

For both B767 and MD11 fleet, it is difficult to find out a trend on the rate and average time for both delays and UGTs. It is possible to conclude that the introduction of the *Reliability Program* and the management decision of giving engineers more responsibility for the resolution of technical problems in 2008, allowed more control over delays and UGTs, mainly for B767 fleet and MD11 UGTs.

Outstations

As a general rule, for both B767 and MD11 the growth of delays and UGTs for Schiphol and worldwide is similar with exception of some particular months. This result is very easy to explain since almost 50% of the departures are from the main station in Amsterdam. The difference between Schiphol and worldwide is mainly for the rate values in both delays and UGTs. The higher rate values in Schiphol are due to the fact that Schiphol is the airline hub and Martinair's most important overhaul station. Sometimes, it is visible that there was a long delay/UGT in the outstations but they were isolated occurrences or incidents in that specific outstation for that particular month. This fact shows that there is no need to make a deeper investigation about any outstations.

1st and 2nd Order of MDCG

Most of delays and UGTs for both B767 and MD11 have been caused by *A/C Mechanical Faults* and *Maintenance* reasons. *Parts* group has also been responsible for a relevant amount of delays. The others MDCG caused randomly delays and UGTs over the years which do not seem to be significant.

The average time of both delays and UGTs were highly dependent of specific events. In general the highest averages were due to a few long delays/UGTs that occurred in that particular quarter and do not seem to be related to any specific MDCG. The average delay time for B767 is slightly lower than for MD11 but both look to be more dependent on *Maintenance* and *A/C Mechanical Fault* groups. *Parts* group is normally responsible for long delays as well. It can be said that, except specific occurrences, *Maintenance, A/C Mechanical Fault* and *Parts* have been responsible for the longest delays and UGTs.

The 2nd Order reason groups revealed to be the same for both B767 and MD11. Therefore, *Replaced LRU* has contributed with around 40% to *A/C Mechanical* delays and UGTs. Other relevant 2nd Order reasons of *A/C Mechanical Fault* are *Replaced Non-LRU* and *Repaired*. Regarding *Maintenance* group, the main reason for the delay has been set as *Deferred/Placard/MEL/HIL.* It is not a surprising result that *Late Out of Maintenance* has also higher contributions since UGTs include, among others situations, extended scheduled maintenance. The *NIL Stock* reason shows to be the most significant one for group *Parts*.

It must not be forgotten that the MDCG are the final causes attributed to the delays/UGTs which do not give enough information to determine the actual root cause. These groups can commonly be used at every MRO or airline and have no specific information about the organization, culture or procedures involved. Another shortcoming of this traditional method used in Martinair is the inability to combine reasons. Many chargeable delays have a non-chargeable contribution to the delay duration.

4. Statistical Analysis

4.1 Experimental Design and Delay Factors

Normally, airlines utilize delay codes which give a good quantitative knowledge on the delays. However, according to Niehues et al [1] and Sridhar [17], it is good practice to apply cause effect diagrams (CAED), also called fishbone diagrams, to identify and understand the root causes and factors. The CAED approach is accomplished with information from people working in the key processes (**Erro! A origem da referência não foi encontrada.**). As this information was not reported, it was necessary to talk directly with the Project Supervisors and Team Leaders. There are numerous factors that can cause a delay and influence the delay time. The ones chosen as independent factors for the statistical analysis were, as follows:

1. Time to A-check (TTA)
2. Aircraft Utilization Rate (AUR)
3. Maintenance Reports (MAREPs)
4. Pilot Reports (PIREPs)
5. Manpower Available (MPA)

The factors above were collected from Martinair database as continuous variables. However, they can be transformed into categorical ones to try to find the best model that fits the data.

TTA: Time to A-check

The Martinair interval between A-checks is set as 700 flight hours for MD11 and 770 flight hours for B767. However, the actual performed date can be delayed by the operator for several reasons. The A-check is divided in 12 parts concerning different tasks (A1-A12). The downtimes of an A-check are normally around 1 day and 7 hours for B767 and 1 day and 3 hours for MD11. It seems reasonable to consider that the interval between A-checks might have an effect on the occurrence of a delay as more *findings*[1] might appear during the normal pre-flight check as the a/c become closer to its A-check. The TTA is calculated as the flight hours left of a particular aircraft to its next A-check.

AUR: Aircraft Utilization Rate

Aircraft utilization is the average daily airborne flying hours or cycles for one aircraft. The aircraft utilization is calculated by the following equation:

$$Aircraf\ Utilization\ Rate\ (AUR) = \frac{Totalflying\ hours\ of\ the\ previous\ 7\ days}{7\ days} \quad (12)$$

AUR represents the average daily utilization of an aircraft for the past 7 days. Normally, MD11 fleet has a higher utilization rate and more cycles, which lead to more maintenance times.

MPA: Manpower Available

In general, the manpower is planned at least one/two weeks in advance. The manpower concerns the number of people to perform maintenance actions on the aircraft (A/c Mechanics). Additionally, there are always one Team Manager and 2-4 Project Supervisors per shift. For the statistical analysis, it was taken the average number of people on duty per shift.

MAREPs: Maintenance Reports

PIREPs: Pilot Reports

All maintenance events, which are carried out on an aircraft, have to be issued in Martinair system as work orders. Sometimes there are unscheduled work orders issued by pilot or maintenance staff. This happens whenever a pilot or an a/c mechanic finds a defect. Each work order corresponds to only one pilot or maintenance complaints (PIREPs and MAREPs, respectively). Even if the complaint is deferred it can cause a delay on that day due to troubleshooting time. It was considered useful to relate the number of

[1] Defect on the aircraft noticed en-route by either flight crew or cabin crew or during a inspection by authorised maintenance personnel

(a)B767

Factors	Description	Value	Unit
TTA	Time to A-check	0 to 835	FH
AUR	Aircraft Utilization	0 to 19h50	FH/day
MPA	Average Manpower available	15 to 46	People / shift
MAREPS	Maintenance Reports Rate	0 to 10	Complaints /Take-off
PIREPS	Pilot Reports Rate	0 to 12	Complaints /Take-off

(b)MD11

Factors	Description	Value	Unit
TTA	Time to A-check	0 to 752	FH
AUR	Aircraft Utilization	0 to 18h37	FH/day
MPA	Average Manpower available	15 to 46	People / shift
MAREPS	Maintenance Reports Rate	0 to 15	Complaints /Take-off
PIREPS	Pilot Reports Rate	0 to 8	Complaints /Take-off

Table 1 - Delay Factors as continuous variables.

Factors	Description	Level	Value	Unit
TTA	Time to A-check	4	0-100, 100-400, 400-600, >600	Flight hours (FH)
AUR	Aircraft Utilization Rate	3	Low (0-7h30), Medium (0-15h), High (>15h)	Flight hours per cycle
MPA	Average Manpower available	4	15-20, 20-30, 30-40, >40	People per shift (/shift)
MAREPS	Maintenance Reports Rate	3	0-2, 2-5, >5	Complaints per Take-off (/TO)
PIREPS	Pilot Reports Rate	3	0-2, 2-5, >5	Complaints per Take-off (/TO)

Table 2 - Delay Factors as categorical variables.

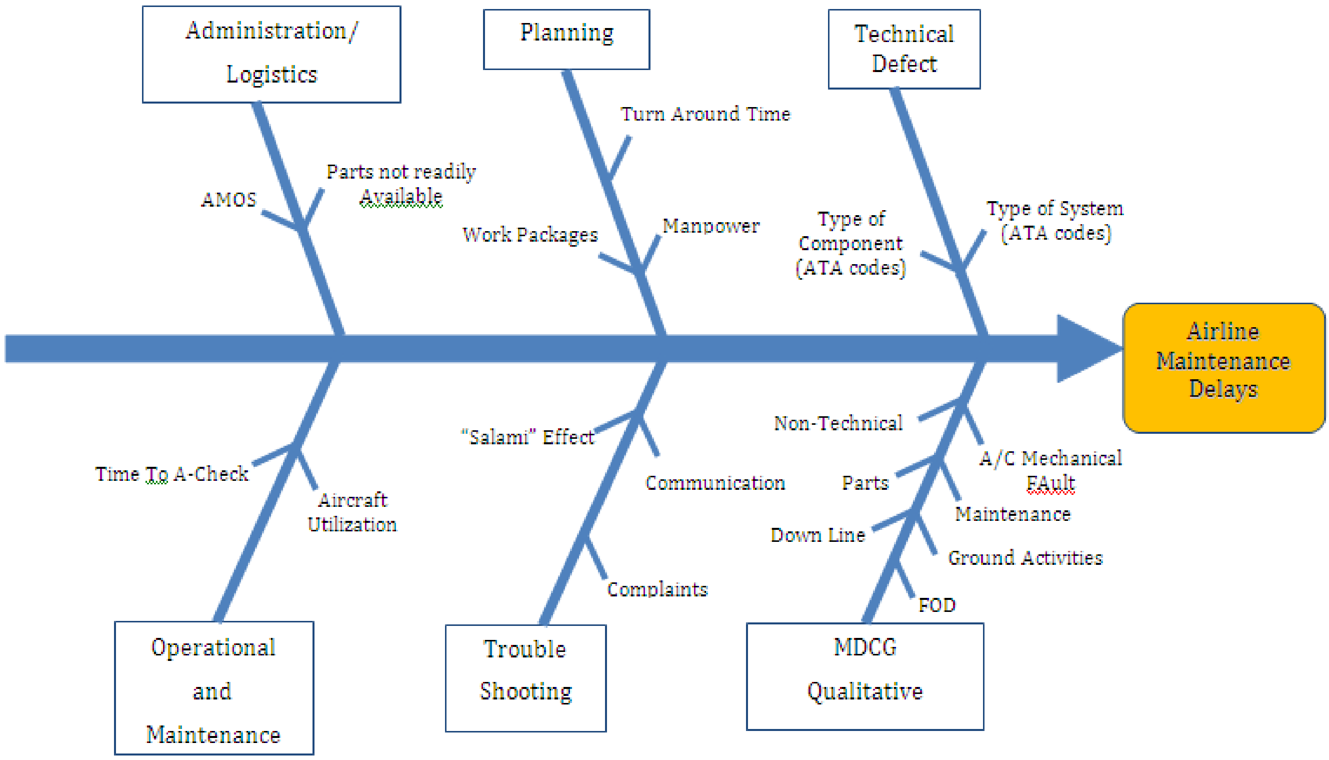

Figure 2- Root Causes for the delays according to experts opinion.

PIREPs and MAREPs per take-off with the delays and UGTs. A MAREPS and PIREPS rate, to use in the statistical analysis, was calculated according to the following equation:

$$MAREPS\ rate = \frac{number\ of\ MAREPS\ on\ a\ day}{number\ of\ take-offs\ on\ the\ same\ day} \quad (13)$$

$$PIREPS\ rate = \frac{number\ of\ PIREPS\ on\ a\ day}{number\ of\ take-offs\ on\ the\ same\ day} \quad (14)$$

These rates take the average number of MAREPs and PIREPs before take-off per day.

For this analysis, we tested 4 different models depending on the type of the aircraft and the event: Delay for B767 fleet, Delay for MD11 fleet, UGT for B767 fleet and UGT for MD11 fleet.

4.2 Experimental Result and Analysis

The first step, it was tested several multiple regression models with a GLM ANOVA analysis to check if there was any relationship between the delay and UGT duration and the response when studied at the same time. Besides some significance was found for some of the factors, the R^2 for all the tested models was below 10%. This low value for R^2 means that the tested model does not fit our data.

The second step was to do a Binary Logistic Regression. Table 3 represents a summary of the results of Binary Logistic Regression for the Delay and UGT as dependent variable for the two types of aircraft. The model for the B767 Delay used predictors as categorical variables while for the other three models they are continuous variables. The Time to A-check (TTA) and the PIREPS rate show significance for all models. However, TTA is significant at the level 0.01 for MD11 models while it is at level 0.05 for B767. The MAREPS rate seems to influence only the UGT models both B767 and MD11. However, for the MD11 Delay model the *p*-value is small (p=0.059). The Manpower Available shows significance at level 0.01 only for the B767 Delay model. One explanation might be the fact that the B767 Delay model is the only one using independent factors as categorical variables. Aircraft Utilization Rate (AUR) is highly significant (p=0.000) for all models except B767 UGT (p=0.821). There are significant second order interactions only for booth MD11 Delay and UGT models (between TTA and the other factors). A significant interaction between factor A and B indicates that the effect of A on the mean value of the dependent variable differs for the various levels of B.

Effect of *TTA*

For MD11 fleet, the coefficient is negative and similar for Delay and UGT models. As Time to A-check increases the probability of a delay and UGT reduces. This means that if the aircraft has come out from an A-check the probability of a delay is lower. Regarding B767, the coefficient of *TTA* is positive for UGT model but it has a relatively small value (0.00092) which can be negligible. For the Delay model, the independent variables are categorical and we need to compare the level coefficients with the reference one (coded as 0). Therefore, it is possible to see a positive effect of *TTA* on the delay. However, the difference between the level coefficients is not very significant.

Effect of *AUR*

For MD11 fleet, the coefficient is negative and similar for Delay and UGT models. As Aircraft Utilization increases the probability of a delay and UGT reduces. If *AUR* is higher the probability of delay and UGT reduces maybe due to more control of the components and system failures and more pre-flight checks. Concerning the B767 Delay model, it seems that there is a higher probability of delays when there is a *Medium* or *High* level of utilization. This means that if an aircraft had had a daily average utilization above 7h30 in the past 7 days, the probability of delay would have increased which wouldn't be surprising. In reality, if an aircraft flies more, the

RESPONSE	DELAY		UGT	
FACTORS **A/C Type**	**B767**	**MD11**	**B767**	**MD11**
TTA	**0.035[b]**	**0.000[a]**	**0.021[5]**	**0.003[1]**
AUR	**0.000[a]**	**0.000[a]**	0.821	**0.000[a]**
MAREPS	0.976	0.059	**0.000[a]**	**0.018[b]**
PIREPS	**0.000[a]**	0.000[1]	**0.001[a]**	**0.000[a]**
MPA	**0.002[a]**	0.603	0.650	0.254
TTA*AUR	0.203	**0.000[a]**	0.154	**0.001[a]**
TTA*MAREPS	0.489	**0.000[a]**	0.130	**0.000[a]**
TTA*PIREPS	0.105	**0.000[a]**	0.461	**0.000[a]**
TTA*MPA	0.546	0.751	0.636	0.442
AUR*MAREPS	0.993	0.997	0.146	0.287
AUR*PIREPS	0.258	0.948	0.155	0.873
AUR*MPA	0.213	0.979	0.482	0.526
MAREPS*PIREPS	0.254	0.105	0.782	0.052
MAREPS*MPA	0.726	0.735	0.947	0.418
PIREPS*MPA	0.968	0.666	0.815	0.840

[a] Significant at the 0.01 level
[b] Significant at the 0.05 level
No superscript denotes lack of significance at both levels (0.01 level and 0.05 level)

Table 3 - A summary of Binary Logistic Regression results for Maintenance Parameters (*p*-values).

		RESPONSE - DELAY
FACTOR	**Level**	**B767**
TTA	0-100	0
	100-400	0.21272
	400-600	0.37024
	600+	0.52755
AUR	Low (0-7h30)	0
	Medium (7h30-15h)	1.91142
	High (15h+)	1.15185
PIREPS	0-2	0
	2-5	0.50051
	5+	1.31110
MPA	15-20	0
	20-30	0.47420
	30-40	0.27380
	40+	1.45672

Table 4 - Coefficients of Fitted Models: Main Effects (categorical variables)

RESPONSE	DELAY	UGT	
FACTORS **A/C Type**	**MD11**	**B767**	**MD11**
TTA	**-0.00389**	**0.00092**	**-0.00391**
AUR	**-2.47471**	-	**-2.66953**
MAREPS	0.08423	**0.510797**	**0.12187**
PIREPS	**0.87697**	**0.161980**	**1.31727**
MPA	-	-	-
TTA*AUR	**0.00885**	-	**0.00805**
MAREPS*TTA	**0.00234**	-	**0.00173**
PIREPS*TTA	**-0.00179**	-	**-0.00151**

Table 5 - Coefficients of Fitted Models: Main Effects (continuous variables).

components and systems have a higher probability of showing malfunctions which can lead to delays.
However, the results show that the probability of delay will be higher if the utilization level is *Medium* instead of *High* (15h+). Once again this result might be due to more maintenance checks (pre-flight checks) for aircrafts with *High* utilization which make possible to be more aware of the problems and planned the corrective actions in a better way.

Effect of *MAREPs rate*

The number of *MAREPs* before take-off is not significant for Delay model both for B767 and MD11. This result is related to lower number of *MAREPs* for delays events. In fact, the delays are mainly caused at the gate (maintenance Pre-flight check and Line Maintenance) where the maintenance actions are not carried out so deeply. For the UGT models the *MAREPs* rate has a positive coefficient, which means a higher number of *MAREPs* gives higher probability of UGTs. This is logical, if A/c Mechanic finds more unexpected defects (reported as *MAREPs*), more time will be needed to troubleshooting and repair which will cause unplanned maintenance (UGTs).

Effect of *PIREPs rate*

The coefficient of *PIREPs* is positive for the 4 models which means that this effect increases the probability of delays and UGTs. In fact, for B767 Delay model, there is a significant increment of the effect of *PIREPs* between the levels. More than 5 *PIREPs* per take-off cause a higher probability of delay. This is logical because before a departure all the complaints (defects) enter in the crew maintenance log have to be performed. For the MD11 the impact of *PIREPs* is higher for the UGT model than for the Delay model. It can be explained by the fact that the crew is responsible to report (*PIREPs*) all defects/alarms noticed during the flight before arrival and if they do not send it at least one hour in advance, there will be a UGT situation.

Effect of MPA

The Manpower available seems to be significant only when used as a categorical variable for the B767 Delay model. The coefficients are always positive which means that if there are more than 20 people (reference level), the probability of delay will increase. It will have more impact when the average number of people per shift is higher than 40 but it will have less impact when there are between 30-40 people available. This result is not compatible with reality because it does not make sense that more people available can cause more delays. It is because the first and the last levels have few observations, probably related to particular days.

5. Predictive Model for Delays and UGTs

5.1 Model Description

The predictive model is developed using regression analysis. This model is able to predict not only the probability of one single event (delay or UGT – unplanned ground times) but also the on-time performance in a certain period.
The user has to decide what he/she wants to simulate and, then he/she has to set the proper options before running the simulation. This model allows the user to choose between:

1. To predict a *Delay* or *UGT*
2. To predict for *B767* or *MD11* fleet
3. To predict the *Probability of one single event* or *On-time Performance*

It must be highlighted that the maintenance delay factors will differ according to the setting chosen. This is explained by the regression analysis results which proved that the significance of the factors is slightly different depending on the type of aircraft and type of event.
Therefore, as stated before, there will be four different regression models: B767 fleet Delays, B767 fleet UGTs, MD11 fleet Delays and MD11 fleet UGTs.

The regression equation and the coefficients of each model differ, but the calculations for the probability of delay/UGT are the same

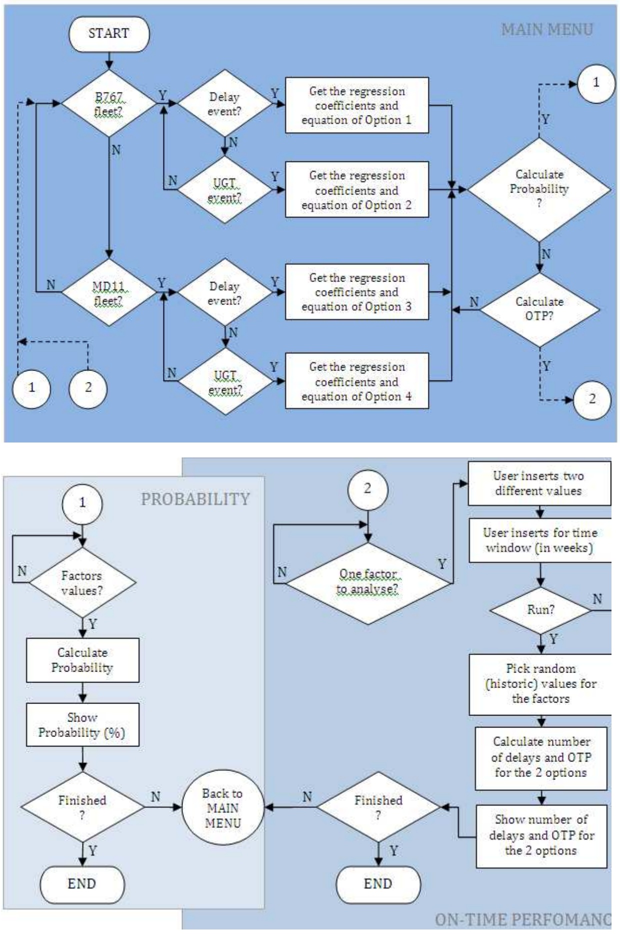

Figure 3 - Model Flowchart

5.2 Simulation

The model can calculate the single probability of one event (delay our UGT) for a certain fleet (B767 or MD11). According to the options selected by the user, the program will ask different inputs (simulation conditions – factors values). The result is shown in percentage and gives the probability of occurrence of a delay/UGT under the selected conditions.

The other tool available is the possibility of predicting the expected number of delays/UGTs and the on-time performance (OTP) for a certain period. In fact, this tool compares the OTP for two different conditions chosen by the user. The objective is to make a fair comparison between different values of one factor, i.e., manager and engineers can predict the impact of variations on that factor. Therefore, the user is asked to choose which factor he/she wants to analyse and to insert two different values for that factor. The user has to define the time window (in weeks)[2]. Then it calculates the probability of delay/UGT for each departure using the proper regression equation. The expected number of delays is basically the sum of all probabilities.

$$END = Expected\ Number\ of\ Delays = \sum_{i=1}^{N} P_i \quad (15)$$

where, N is the total number of departures and P_i is probability of delay for each departure.

The On-Time Performance (OTP) is given by the equation 16.

$$OTP = \left(1 - \frac{Expected\ number\ of\ delays}{Total\ number\ of\ departures}\right) x100 \quad (16)$$

Illustrative Example

Some simulations were made to predict the OTP for B767 and MD11 fleet. The time window for all simulations was 4 weeks (1 month). The average number of flights per week for the B767 fleet is 50 and for the MD11 is 130. Thus, these simulations took around 25 seconds to show the results for the B767 fleet and 55 seconds for the MD11 fleet. These simulations allow us to compare the impact of the different factors, i.e., how each factor influences the on-time performance and the number of delays/UGTs. The Table 6 shows some of these examples.

Model Validation and Standard Error

Cross-validation is mainly used in settings where the goal is prediction, and we want to estimate how accurately a predictive model will perform in practice Moreover, cross-validation can be used to compare the performances of different predictive models. As suggested in the literature, the sample data was divided into a *training set* to perform the analysis

[2] The model assumes that in one week there are in average 50 departures for B767 fleet and 130 for MD11.

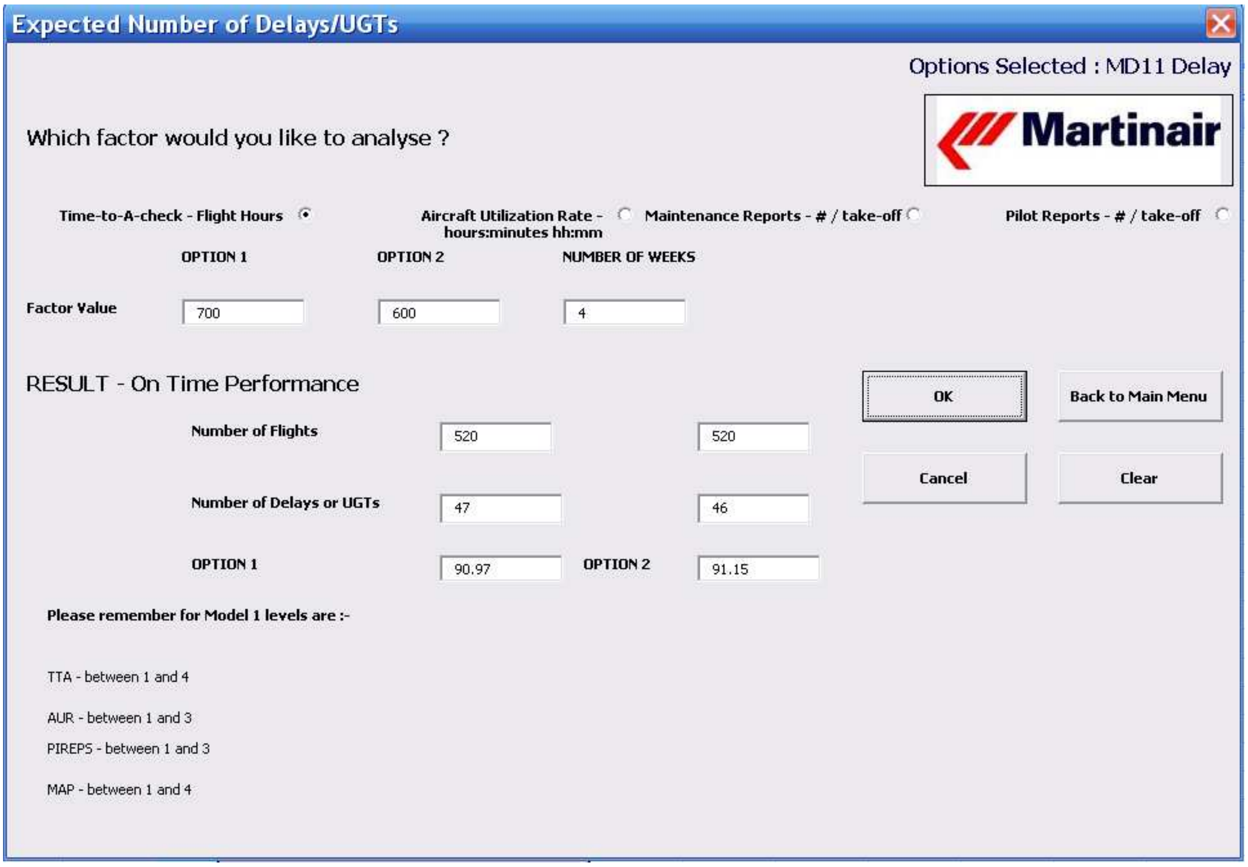

Figure 4 - Example of the dialog box for the simulation of the On-Time Performance

B767 Delays	**TTA**		**AUR**		**PIREPS**		**MPA**	
Simulations / **Results**	Level 3 400-600 FH	Level 4 > 600 FH	Level 2 7:30-15:00 FH	Level 3 >15:00 FH	Level 1 0-2 /TO	Level 2 2-5 /TO	Level 2 20-30 people	Level 3 30-40 people
Number of Delays	14	15	17	9	10	17	14	11
OTP (%)	92.8	92.4	91.4	95.5	94.9	91.5	93.2	94.5

MD11 Delays	**TTA**		**AUR**		**MAREPS**		**PIREPS**	
Simulations / **Results**	550 FH	700 FH	08:00 FH	15:30 FH	1 /TO	2 /TO	1 /TO	2 /TO
Number of Delays	34	32	31	34	54	116	38	57
OTP (%)	93.5	93.8	94.1	93.5	89.5	77.8	92.7	89.1

Table 6 - Results of the simulation for the expected number of delays for one month period (200 flights for B767 fleet and 520 flights for MD11).

(90% of the sample) and a *testing/validating set* to validate the analysis (other 10% of the sample).

To reduce variability, multiple rounds of cross-validation are performed using different partitions and the validation results are averaged over the rounds. The cross-validation was used in two different ways: to choose the best model for each option and to calculate the standard error.

The log-likelihood function was used to check how good the model in test was performing. The statistical result proves it is possible to use these models for forecasting. The regression model that presents the highest value for log-likelihood (closest number to 0) was chosen.

The standard error results for the different options are presented in Table 7. It was simulated 50 flights for each option and the simulation was repeated 1000 times.

OPTION	Number of Delays/UGTs		On-Time Performance	
	Mean	**Standard Error**	Mean	**Standard Error**
1. B767 Delay	4	**± 1.5**	92%	**± 3.0%**
2. B767 UGT	1.5	**± 1.1**	97%	**± 2.2%**
3. MD11 Delay	3	**± 1.6**	94%	**± 3.2%**
4. MD11 UGT	2	**± 1.2**	96%	**± 2.4%**

Table 7 - Mean and Standard Deviation for number of delays and OTP

6. Conclusions and Recommendations

6.1 Statistical Analysis

The maintenance factors chosen for the statistical analysis were Time-to A-check, Aircraft Utilization Rate, Maintenance and Pilot Reports and Manpower Available. Even if there are other important factors that can influence the occurrence and time of a delay, they were not selected because there was not any historical data available for a statistical analysis. Among these factors are Troubleshooting, Administration, Logistic and Planning. Using a statistical software package it was possible to find out some conclusions about delays and UGTs.

- Trying to find a correlation between the maintenance factors selected and the duration of delays and UGTs proved to be useless. No correlation was found between the independent variables (factors) and the dependent one (delay time).
- The factors proved to be significant when using a Binary Logistic Regression to predict the probability of delay and UGT.

6.2 The Predictive Model

The predictive model is developed using regression analysis and is able to predict the probability of one single event (delay or UGT – unplanned ground times) and also the on-time performance (OTP) in a certain period (in weeks). The model is validated using cross-validation and the log-likelihood function. The standard error was also calculated for the tool that calculates the number of delays and OTP. It was simulated 50 flights for each of the four options and the simulation was repeated 1000 times. The results showed that the standard error is around 1.5 in 3.5 expected delays and 3.1% in 93% for the on-time performance.

The probabilities of delay and UGTs are, as expected, always low values. For the M&E department this information might not be very useful. For example, knowing that the delay probability of a single flight is 10% instead of 6% may not be a good enough reason for taking measures to avoid the delay because most definitely it would imply some costs.

A powerful tool is to estimate the number of delays and the on-time performance for the next week, month or year using the maintenance factors. The model is able to predict how good Martinair's fleet (B767 or MD11 fleet) will perform for the desire period. The tool also allows engineers and managers to evaluate how changes in the *maintenance factors* will impact Martinair's performance in a long term period.

Some simulation results showed that *Time to A-check (TTA)* for both B767 and MD11 fleet do not have much influence in the OTP which suggests that there is no benefit in changing the A-checks intervals. The *Maintenance and Pilot Reports (MAREPS and PIREPS)* are, as expected, highly correlated to the number of delays for both B767 and MD11 fleet. However, it is difficult to control these factors, anyway if M&E estimates the rate of PIREPS and MAREPS before take-off, it is able to control better its performance. The *Aircraft Utilization Rate (AUR)* seems to have some impact only for the B767 fleet increasing the OTP by 4.1% when the Aircraft Utilization rate is higher than 15 flight hours per day. This result might be justified by the fact that an aircraft that flies more it is subject to more maintenance inspections (pre-flight checks). On one hand more inspections might help in controlling the technical problems better, on the other hand the technical problems might be easily deferred to the next landing. For the MD11 fleet, a higher AUR affects the OTP by 1% which cannot be taking into account because of the associated standard error.

The maintenance factors chosen revealed to be statistical significant and allow a valid prediction of the number of delays and OTP. However, the information that can be extracted for a single factor might not be good enough for M&E to take actions on that maintenance factor and control the number of delays.

6.3 Recommendations and Future Work

The following suggestions are made to Martinair M&E.

- It is important to understand where is M&E wasting time when trying to solve a defect that can lead to delay. According to expert opinion, delays are often caused by troubleshooting. However, this information it is not automatically reported. It would be a good idea to try to report this issue in AMOS system or to do an intensive investigation in the hangar 32 and at the gate in Schiphol to be able to draw some conclusions.
- As *Replaced LRU* was one of the mainly reasons for the cause of delays, it might be interesting to study the *No-Fault Found*, i.e., sometimes a Rotable component was suspected to be failed and removed from aircraft for that reason, but during testing, it is found to be in a serviceable condition.
- As the *Aircraft Aging* is another factor for the delays and unplanned ground times, it should be considered another research on this single topic.
- In order to get better predictions of the on-time performance and number of delays, it might be good idea to do a similar statistical analysis using technical maintenance factors, as the ATA chapters. This was not developed in this research because we were looking for non-chargeable delays.
- Martinair should find out how much it costs one minute of delay to be able to perform cost benefit analysis regarding the delays and maintenance factors.

7. References

[1] Niehues, A., Belin, S., Hansson, T., Hauser, R., Mostajo, M. and J. Ritcher "Punctuality: How Airlines Can Improve On-Time Performance", Booz-Allen & Hamilton (2001)

[2] Yuan, D. "Flight Delay-Cost Simulation and Airline Schedule Optimization", Faculty of Science, Engineering and Technology RMIT University, Victoria, Australia (2007)

[3] Knotts, R. M. H. "Civil aircraft maintenance and support: A fault diagnosis from a business perspective", *Journal of Quality in Maintenance*

Engineering, Vol. 5, No. 4, pages 335–347 (1999)

[4] Sridhar, K. “Delays – How to Mitigate It” *Maintenance Reliability & Cost analysis Seminar,* Boeing, Section 18 (2008)

[5] Australian Transport Safety Bureau (ATSB) “How old is too old? The impact of aging aircraft on flight safety” Aviation Research and Analysis Report – B20050205 (2007)

[6] AhmadBeygi,S., Cohn A. and M. Lapp “Decreasing Airline Delay Propagation by Re-Allocating Scheduled Slack” *Annual Conference*, Alfred P. Sloan Foundation Industry Studies Program (2008)

[7] AhmadBeygi, S., Cohn A., Guan, Y. And P. Belobaba “Analysis of the Potential Delay Propagation in Passenger Airline Network” *Journal of Air Transport Management*, Vol.14, pages 221-236 (2008)

[8] Abdelghany K., Sharmila S., Raina S. And A. Abdelghany “A Model for Projecting Flight Delays during Irregular Operation Conditions” *Journal of Air Transport Management,* Vol. 10 pages 385-394 (2004)

[9] Wu, C. L. And R. E. Caves “Towards the optimisation of the Scheduled Reliability of Aircraft Rotations” *Journal of Air Transport Management*, Vol.8, pages 419-426 (2002)

[10] Wu, C. L. And R. E. Caves “Aircraft Operational Costs and Turnaround Efficiency of Airports” *Journal of Air Transport Management*, Vol.6, pages 201-208 (2000)

[11] Wu, C. L. And R. E. Caves “Inherent Delays and Operational Reliability of Airline Schedules” *Journal of Air Transport Management*, Vol.11, pages 273-282 (2005)

[12] Yufeng, T., Ball M.O. and W.S. Jank “Estimating Flight Departure Delays Distributions – A Statistical Approach With Long-Term Trend and Short-Term Pattern” *Journal of America Statistical Association*, Vol.103, pages 112-125 (2008)

[13] Mueller, E.R. and G.B. Chatteji “Analysis of Aircraft Arrival and Departure Delay Characteristics” *Aircraft Technology, Integration and Operations* America Institute of Aeronautics and Astronautics, AIAA 2002-5866 (2002)

[14] Sachon, M. and E. Pate “Delay and Safety in Airline Maintenance” *Reliability engineering & System Safety*, Vol. 67, pages 301-309 (2000)

[15] Sridhar, K. “Delays – How to Mitigate It” *Maintenance Reliability & Cost analysis Seminar,* Boeing, Section 18 (2008)

[16] Sridhar, K. “Data Collection System” *Maintenance Reliability & Cost analysis Seminar*, Boeing, Section 4 (2008)

[17] Sridhar, K. “Data Analysis and Corrective Action” *Maintenance Reliability & Cost analysis Seminar,* Boeing, Section 6 (2008)

[18] Freund, R. And W. Wilson “Regression Analysis: Statistical Modelling of a Response Variable”, Academic Press, Oxford, UK (1998)

[19] Miles, J. and M. Shevlin “Applying Regression and Correlation: a guide for students and researchers”, Sage Publications, London (2001)

[20] Guerard, J. and E. Schwartz “Quantitative Corporate Finance”, Springer Verlag, New Yorker, USA (2007)

[21] Ghobbar, A. and C. Friend “Evaluation of Forecasting Methods for Intermittent Parts Demand in the Field of Aviation: a Predictive Model”, *Computer and Operations Research*, Vol. 30, pages 2097-2114 (2003)

[22] Montgomery, D. C. and G. C. Runger “Applied Statistics and Probability for Engineers”, John Wiley and Sons, New York, USA (2003)

[23] Minitab Inc. Minitab Statistical Software, Release 15 for Windows, State College, Pennsylvania (2006)

[24] Weisber, S. “Applied Linear Regression”, Wiley-Interscience, New Jersey, USA (2005)

[25] Dekking, F., Kraaikamo C., Lopuhaä, H. and L. Meester “A Modern Introduction to Probability and Statistics – Understanding Why and How”, Springer, London, UK (2005)

[26] Kohavi, R. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”, *Proceeding of the Fourteenth International Joint Conference on Artificial Intelligence*, Vol. 2 (12), pages 1137-1143 (1995)