A DSS for business decisions in air transportation: a case study

Original

Availability:
This version is available at: 11583/2438778 since:

Publisher:
IEEE

Published
DOI:10.1109/ISDA.2011.6121749

Terms of use:
openAccess
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
ieee

copyright 20xx IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating .

(Article begins on next page)
Abstract—The socio-economic development leads people to a great mobility. Thus the flights identification and management is becoming a key factor for the economic growth of the areas nearby the airports. The airport management is constantly looking for methods to improve its performance, both in terms of profitability and quality of service and the proper planning of passenger flows. To address these issues, scientific research provides methods and tools for decision support at all planning levels (i.e., strategic, tactical, operational, real time). In recent literature, it is now widely recognized that the hybridization of simulation and optimization systems is a very reliable technique for such decisions. This work intends to present an efficient Decision Support System framework based on the hybridization of a discrete event simulator and a Logit model. In order to show the effectiveness of the framework, we show the results of a real case study in North Italy.

Keywords—Logit model; Simulation; Optimization; Decision Support System.

I. INTRODUCTION

In air transportation, the decisions on which flights should be opened for an airport consists of studying demand, supply and their economic and spatial interrelationships. Many transformations are taking place in the air transportation system, due to the presence of different airlines and airports characterized by different costs and service levels, and in strong competition each other. In order to understand and control this important phenomenon, it is necessary to be able to model the competition in the system and measure the impact of transformations, as well as to give to the air system management new methods for dealing with decisions inherent the potential increase in profit and traffic due to changes in their offers. In the literature most of the existing models devoted to similar problems are multinomial Logit models, derived from spatial interaction and discrete choice theories ([5], [11], [12], [3]). Notwithstanding their solid theoretical structure, these models present some behavioural and structural anomalies, like the Independence from Irrelevant Alternatives (IIA) and the Regularity properties, which make them unsuitable for a correct impact analysis. Several applications of pure multinomial Logit models can be found in location [9] and transportation [10]. In air transportation systems, we recall the passenger choice in San Francisco Bay by Hess and Polak [3] and the results of Coldren et al. in predicting aggregate flows between different cities [2]. In this paper we introduce an efficient Logit formulation for modeling the competition between different airports and forecasting the effects of the introduction of modifications of the flight schedule in the system. In order to deal with the stochastic sources not taken into account by the Logit model, it is integrated with a simulation framework. An application of the developed framework to a real Italian airport is presented.

The paper is organized as follows. Section II gives a better insight on the problem itself. The overall scheme of the DSS as well as the Logit model and an application of the framework are presented in Section III and IV, respectively. Finally Section V presents the final remarks and future developments.

II. PROBLEM DEFINITION AND LITERATURE REVIEW

Let us consider a set of airports for which we know the flights connecting each other, as well as the flights schedule and the overall passenger flow of each flight connection. Our aim is to predict the flows due to a frequency change of an existing flight or to the opening of a new one between a given pair of airports.

The passengers in the catchment area of each airport are unknown, but some forecasts on this information are given. Then, the passenger number becomes a stochastic variable with unknown probability distribution. For each flight we know in advance the mean cost paid by a single passenger for taking the flight, as well as the quality parameters of the airports, including the time needed to reach the airport from the nearby area and the time needed to perform internal operations such as passport checks and luggage delivery.

The optimization methods described in literature focus on one of the actors involved in the airport system (users, airport managers, government agencies and airlines), losing the overall view that the problem requires, or simplifying the complex cause-effect relationships between decision makers, in particular the behavior of passengers. Therefore, the base of this work is an economic and spatial interaction Logit model aimed at modeling the dependency between the actors. This kind of model shows a high adaptability to changing decision-making levers involved and a wide
efficiency demonstrated in several fields, including transport and retail [4]. Further, Logit models implicitly consider some sources of uncertainty due to the choice of the passengers, while do not consider other uncertainty sources, like demand or offer forecasts. In the airport market the most frequent errors are the estimates of the passengers to various destinations and their splitting into sub-categories, as well as the uncertainty about the demand in terms of passengers for each airport flight. The difficulty of modeling requires the introduction of analytical tools to evaluate the entire system and suggest corrective actions, in other words, there may be many possible realizations of the parameters, each of which is a possible scenario.

Scientific research has recently focused on developing models that combine simulation and optimization. The simulator carries out the generation of possible scenarios of the system under analysis to evaluate the decisions taken by the model and predict their impact on the overall system. The optimization block finds the optimal solution for each scenario starting from the parameters identified by the simulation and solving a combinatorial optimization problem. The hybridization of the two systems allows to verify and refine the decisions taken. In particular, Simulation and Optimization methods allow to incorporate into the model the dynamic component of the system due to structural changes and to evaluate the effect of decisions over the medium term. Moreover, they are effective in finding a solution that is feasible for all the considered scenarios, and minimizing the deviation of the overall solution from the optimal solution for each scenario. Although some applications combine simulation and optimization using linear model, there is still significant potential for improvement ([1], [8]). The main direction that we investigate in this work is the introduction of non-linear potential for improvement ([1], [8]). The main direction that we investigate in this work is the introduction of non-linear potential for improvement ([1], [8]).

III. THE DSS FRAMEWORK

The framework, depicted in Figure 1, is composed by a simulator which, given the distribution of the total supply of each airport and how it is split among the existing flights of the airport, generates a series of scenarios. Each scenario is then used by a Logit model to calibrate the flow matrix (step 1), to simulate a given change in this matrix and to predict the new passenger flows. The output of each scenario is then used by the simulator in order to combine it with the other ones (step 2) and perform a statistical analysis on the aggregate results (step 3). In order to make a more accurate definition of the travel times and costs matrices, a georeference module is used. The georeference module is implemented by means of Google Earth APIs and it is also used to graphically represents the results of the DSS itself. Finally, a post-optimization software module is devoted to choose the best features of the new flight (e.g. the kind of aircraft, the number of flight).

The framework has been implemented in C++ and Math Kernel Library (MKL) to maximize the performances [7]. For the simulation phase, it has been implemented by means of Omnet++ 4.1 [6].

In the following we focus our discussion on the Logit model in order to give an highlight of the optimization core of the framework.

A. Logit model

Throughout this paper, \( x \), \( x \) and \( X \) denote a generic scalar, vector (lowercase and boldface), and matrix (uppercase and boldface), respectively. Superscript \( \sim \) and \( \hat{\sim} \) will stand for the observed and estimated values, i.e. the values given to the Logit as input by simulator and the flows obtained after a parameter calibration, respectively.

Given a set of \( n \) origin airports, a set of \( m \) destination airports and \( n \times m \) flights connecting them, we define

- identifier of origin airport \( i \in \{1, \ldots, n\} \);
- identifier of destination airport \( j \in \{1, \ldots, m\} \);
- observed flows matrix \( \hat{T} \in \mathbb{R}^{n \times m} \); i.e. elements \( \hat{T}_{i,j} \) give the number of passengers depart between \( i \) and \( j \).
- generalized travel cost matrix \( C \in \mathbb{R}^{n \times m} \), i.e. elements \( C_{i,j} \) give the travel cost of the flight arriving in \( j \) and departing from the airport \( i \).
- flight frequency matrix \( V \in \mathbb{R}^{n \times m} \), i.e. elements \( V_{i,j} \) give the number of flight from \( i \) to \( j \) in a fixed time period (usually, the weekly schedule).

We also define the total supply vector \( \mathbf{o} \in \mathbb{R}^{n} \) in passenger of origin airports (1) and the total demand vector \( \mathbf{d} \in \mathbb{R}^{m} \) of flights (2).

\[
\mathbf{o} = \sum_{j} \hat{T}_{i,j} \quad \forall i \\
\mathbf{d} = \sum_{i} \hat{T}_{i,j} \quad \forall j
\]

Our aim is to define a proper model for the estimated flows matrix \( \hat{T} \) considering both the operational characteristics of the flights connecting \( i \) and \( j \), as well as the peculiar features of the destination airports. This model must be able to reproduce the matrix of the observed flows \( \mathbf{T} \). To do that we develop a multinomial Logit model, which generates...
the estimated flows matrix, as well as the main parameters characterizing the system.

The Logit model is represented in (3) where \( w \in \mathbb{R}^m \) is an attraction factor of the flight destination \( j \), \( \beta \) is a distance decay parameter and \( \gamma \) is the flight frequency decay parameter.

\[
\tilde{T}'_{ij} = \frac{o_{ij} \cdot e^{-\beta \tilde{C}_{ij}} \cdot e^{\gamma \tilde{V}_{ij}}}{\sum_j o_j \cdot e^{-\beta \tilde{C}_{ij}} \cdot e^{\gamma \tilde{V}_{ij}}}
\]

In order to predict a change in the system due to a new flight or a schedule change, first we have to find the values of \( w \), \( \beta \) and \( \gamma \) characterizing the system and such that they reproduce the observed flows \( \tilde{T} \). This is done in the following calibration phase. The procedure works as follows:

- Initialization. Set \( \gamma_0 = \frac{2}{3}, \beta_0 = \frac{2}{3} \)
- While the values of \( \gamma \) and \( \beta \), and \( w \) changes over a given threshold or a maximum number of iterations is not reached
  - Given \( \gamma_{k-1} \) and \( \beta_{k-1} \), find the values of \( w_k \) which are the roots of the system
    \[
d - \sum_i \tilde{T}_{i,j} = 0 \quad \forall j
    \]
  - Given \( w_k \) and \( \beta_{k-1} \), find the new value of \( \gamma_k \) by considering which is the root of the system
    \[
    \sum_{i,j} (\tilde{V} - \bar{v}) \tilde{T} = 0
    \]
  - Given \( w_k \) and \( \gamma_k \), find the new value of \( \beta_k \) which is the root of the system
    \[
    \sum_{i,j} (\tilde{C} - \bar{c}) \tilde{T} = 0
    \]
- Find a correlation between \( w \) and the economical and structural features of the corresponding airports, indicated respectively \( \bar{\rho} \in \mathbb{R}^m \) and \( \bar{\sigma} \in \mathbb{R}^m \). The economical and structural features we consider are the total number of flights departing from an airport, the service time of the passengers, obtained as the sum of the time spent by a passenger in the airport before being his boarding and the time needed to access the airport by ground services as taxi, private car and train. This is done by a logarithmic transformation of the attraction factors estimated through linear regression. The correlation between \( w \) and two regression coefficients \( \alpha \) and \( \delta \) is shown in (7).

\[
\ln(w) = \alpha [\bar{\rho}] + \delta
\]

Equations (4) ensure that the estimated flows generated by the origin airports are equal to their demand, while (5) and (6) force the sum of the estimated flows weighted by \( \tilde{V} \) and \( \tilde{C} \) to be equal to the mean number of flight and cost, respectively.

The calibration of \( w \), \( \beta \), and \( \gamma \) is implemented by means of a fixed point algorithm, while the values of \( \alpha \) and \( \delta \) are obtained by a linear regression. Notice that the procedure of system parameters estimation usually stops when the values of the parameters themselves at iteration \( k \) do not differ from the values of iteration \( k - 1 \) more than a given threshold, set to \( 10^{-3} \). A second stopping criterion on the maximum number of iterations is given in order to prevent numerical issues.

By substituting equation (7) in (3), the new expression of the Logit model is

\[
\tilde{T}'_{ij} = \frac{o_{ij} \cdot e^{\alpha \hat{\rho}_j + \alpha_2 \hat{\sigma}_j + \delta} \cdot e^{-\beta \hat{C}_{ij}} \cdot e^{\gamma \hat{V}_{ij}}}{\sum_j o_j \cdot e^{\alpha \hat{\rho}_j + \alpha_2 \hat{\sigma}_j + \delta} \cdot e^{-\beta \hat{C}_{ij}} \cdot e^{\gamma \hat{V}_{ij}}}.
\]

Even if the parameters have been accurately calibrated, they cannot take into account all the economical and structural features of the system. Thus, these factors are considered by a perturbation matrix \( \Theta \in \mathbb{R}^{n \times m} \). This matrix is defined by the ratio of observed flow in passenger (\( \hat{T} \)) to computed flow with Logit model (\( T \)).

\[
\Theta_{ij} = \frac{\hat{T}_{ij}}{T_{ij}} \quad \forall i, j
\]

Each \( \Theta_{ij} \) measures in percentage the impact of these unknown factors on the flow (\( i, j \)). Thus, the full expression of the estimated flows becomes

\[
\tilde{T}_{ij} = \Theta_{ij} \cdot o_{ij} \cdot e^{\alpha \hat{\rho}_j + \alpha_2 \hat{\sigma}_j + \delta} \cdot e^{-\beta \hat{C}_{ij}} \cdot e^{\gamma \hat{V}_{ij}}.
\]

### B. Simulator

Given the Logit model, we can compute the new passenger flows by introducing the changes (e.g. the opening of a new flight or a flight cost change). To do this, we developed a modular simulator that is formed by a generator of scenarios, an optimization block that implements the Logit model and a set of data analysis tools. The scenario-based simulation reduces the probability of error on the results and considers the changes in the flow of passengers over time. The scenarios are created by random modifications, with chosen probability distribution, of observed flows \( \hat{T} \). The number of scenarios is defined ad hoc by the relative standard deviation (RSD) of data set and must ensure that the solution is stable and statistically independent of the simulation. A typical value of the error on the results of the simulation is 2%.

Once model has been calibrated on the scenario by the optimization block, the simulation process needs some input parameters that describe the changes in the system in terms of flight schedules and flight costs. The changes of the flows are calculated as the difference between the flows \( \hat{T} \) obtained after introducing the new flights schedules and observed flows \( T \). This gives the new catchment area of
the airports in terms of passengers, as well as a forecasts of the flows on each flight. The analysis block stores the results of the variation for each scenario and computes some statistical analysis on the new flows distributions. Finally, a post optimization process checks the solution that maximizes the effect on the airport of origin and chooses the parameters of the selected flight, i.e. flight frequency, airplane size.

Notice that the model can evaluate more than one new opening at a time and it is interesting to observe that it also allows to simulate reaction policies from the existing airports by assigning different values to their economical and structural features (e.g. an existing airport could decide to reduce its prices in order to face the competition of the new flight).

IV. COMPUTATIONAL TEST: THE BOLZANO AIRPORT CASE STUDY

In order to validate the framework, we simulate the consequences of opening new flights in Bolzano airport (in the following ABD). We considered the case of opening new flights between a panel of given destinations, selected from the database of International Air Transport Association (IATA).

The catchment area generates a stochastic supply, which is considered to be uniformly distributed, while the other data (costs, flight frequency, access times) have been taken from public documents of the airports themselves and the databases of Eurostat and IATA. The forecasts generated by the framework have been validated by BDS s.r.l., a consulting company and specialized staff of Bolzano airport.

A. Bolzano - Dolomiti

Bolzano Dolomiti Airport (ABD) was born in 1992 and the work of modernization of the structure was completed in 1999. The offer of flights provided by ABD Airport is very limited and the only possible flight is between Bolzano and Rome with daily frequency of four times per day (two in the morning and two in the afternoon).

We want to predict the opening of a new route to manage the flow of passengers from the airport to a chosen destination. This type of problem is called out-bound. The simulations performed on ABD Airport aim to establish the route, the type of aircraft, the flight frequency and type of airline (business or low cost) that allows to attract the highest number of passengers.

Settings of the simulator: The catchment area is focused on the residents of Bolzano area moving in neighboring countries for business and tourists who come from all over Europe for the summer or winter holidays. Analyzing the socioeconomic conditions of Bolzano and some statistic surveys on data supplied by the provincial statistic institute (ASTAP) and on samples of users of ABD Airport, the cluster of destinations and the set of competitors have been defined. The identified macro-areas are Germany, Austria and Switzerland and the respective centroid are Frankfurt, Vienna, Zurich. Similarly, we defined the cluster of sources such as airports easily reachable from Bolzano. In particular, competitors are Venice, Verona, Treviso, Innsbruck, Salzburg and Monaco.

For each route, we define the possible price ranges and the flight frequency. The cost is considered excluding airport taxes and typical ranges are:

- Low: cost lower than 200 €.
- Medium: cost ranges between 200 € and 300 €.
- High: cost ranges between 300 € and 500 €.

According to the usual schedules of ABD Airport, the simulation considers a low frequency (two aircraft per day) and a high frequency (four aircraft per day). Calibrating the number of scenarios, a stable solution is obtained by simulating 30 scenarios.

Simulation phase: The simulation results are summarized in tables IV-A and IV-A (first two rows) and in Figure 2. In particular, we show the outgoing flows from ABD Airport grouped by price ranges and by flight frequency in table IV-A and the simulated flows of competitors due to the opening of the flight ABD Airport - Frankfurt in table IV-A.

The Bolzano-Frankfurt route provides the largest flows area for both frequency values and for different ticket price range. The validity of the solution is demonstrated by data on tourism of Bolzano, which certify the presence of a large German component on the territory. The optimal setting is given by an high frequency flight schedule with a low-cost airline.

<table>
<thead>
<tr>
<th>Destination</th>
<th>Flights per day</th>
<th>T&lt;sub&gt;200&lt;/sub&gt;</th>
<th>T&lt;sub&gt;300&lt;/sub&gt;</th>
<th>T&lt;sub&gt;500&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frankfurt</td>
<td>2</td>
<td>5968</td>
<td>5140</td>
<td>4426</td>
</tr>
<tr>
<td>Frankfurt</td>
<td>4</td>
<td>32504</td>
<td>28078</td>
<td>24243</td>
</tr>
<tr>
<td>Wien</td>
<td>2</td>
<td>3833</td>
<td>3302</td>
<td>2844</td>
</tr>
<tr>
<td>Wien</td>
<td>4</td>
<td>19154</td>
<td>16553</td>
<td>14297</td>
</tr>
<tr>
<td>Zurich</td>
<td>2</td>
<td>2205</td>
<td>1890</td>
<td>1636</td>
</tr>
<tr>
<td>Zurich</td>
<td>4</td>
<td>12272</td>
<td>10605</td>
<td>9160</td>
</tr>
</tbody>
</table>

Figure 2. Results of the three simulated flights with four daily flights. A bar chart shows the intercepted flow by each route. The bands of low, medium and high cost are indicated by green, yellow and orange bars, respectively (example of georeferencing generated by the framework).
Verona airport is the competitor which has contributed most to the flow of new offer. The following section examines the involvement of Verona to address the reduction in passengers.

**Reaction phase:** The simulation of a reaction of an airport is a long-term analysis and it assumes that the new route has become a reality. One possible operation is to reduce the cost of the ticket that is more affected by the new system or the transition to a low-cost airline.

We have created two scenarios. The first reduces the price of the route Verona-Frankfurt to a Medium range. The second reduces further the cost to a Low range of costs and requires some capital investment for the research and the opening of a low-cost airline and an average aircraft load factor equal to 72%. Both simulations allow to attract part of the catchment area. In particular, the introduction of a low-cost airline offers a number of passengers greater than the initial one. Flows of the entire system are shown in the last two rows of Table IV-A and in Figure 3.

The price reduction affects all of the competitor airports. In more details, while the losses are marginal when Verona changes its price to a medium range, it becomes more relevant when a low-cost flight is introduced, especially for Munich ($\Delta T_L = -12601$), Salzburg ($\Delta T_L = -3836$) e Venice ($\Delta T_L = -2805$). Their losses are negligible when compared to the total flow of each airport and will hardly lead to a subsequent reaction step. Similarly, the transition to a low-cost does not affect the number of passengers using the route Bolzano - Frankfurt, which can coexist with other routes of the analyzed system.

**Additional transport costs:** So far, the simulations performed did not consider the costs of transportation to the airport of origin. Analyzing the situation of Bolzano we must consider the real cost that a user has to support to reach a competitor airport.

The total cost of a route is defined in the equation (11), where $C_{air}$ is the cost of air ticket, $C_{int mode}$ is the estimated cost due to the intermodal network connecting the airport with the surrounding area and $K$ and $t_{car}$ represent constants that quantify the time spent to move in euro and the travel time by car from Bolzano, respectively.

$$C_{tot} = C_{air} + C_{int mode} + K \cdot t_{car} \quad \forall j \quad (11)$$

Simulating the system with the same price ranges and the same frequencies of the flight, we achieved the results in table IV-A. Even in this case, the route to Frankfurt provides a greater catchment area if it is associated with a low-cost airline and a high frequency of flights. The increasing of simulated flow is due to the introduction of additional costs; the decision of the users is also characterized by a spatial factor (i.e. the distance from the origin).

About 80% of attracted passengers by the new route comes from Verona, which will establish the reaction policies (i.e. reduction of the ticket price and the transition to a low-cost airline).

In Table IV-A we present the results of the simulation of a reduction of ticket price by Verona in order to be more competitive on the flight to Frankfurt. The simulation results show that simply reducing the cost from High to Medium is not enough and implies a further loss of passengers of $\Delta T_M = -3511$. The introduction of a low-cost flight ensures an increase in the catchment area, however, the final flow is reduced by about 15% compared with the observed values.

The route Bolzano - Frankfurt is not influenced by the reaction policies of Verona. The variation of the flow is less than 1%, given that the survival of the route is guaranteed.

### V. Conclusion

In this paper we presented a simulation-optimization framework for the optimization of flight connections. The framework is able to consider both stochasticity due to the uncertainty of the data and to represent the choices due to the reaction of customers to a change in the airport system under study.

A real test on airport of Bolzano shows how the framework is able to forecast the new flows due to the opening of a flight connection, as well as to take into account the reaction of competitors due to this opening.
From a computational point of view, the forecast can be obtained with a negligible computational effort (less than a minute for a full run of the simulation with 5 competitors over a total of 15 flight destinations).

In the near future we aim to incorporate other variants of Logit models. In particular we are interested in the variants considering multi-echelon networks in order to model hub flows.

ACKNOWLEDGMENT

The authors want to thank BDS Consulting s.r.l. and the airport of Bolzano for their support. While working of this paper, Gabriele Benedetti was supported by Fondazione ISI under its Lagrange Project Industrial Grants.

REFERENCES


