Worldwide Passenger Flows Estimation

Rodrigo Acuna-Agost[1], Zacharie ALES[2], Ezequiel Geremia[1], Thiago Gouveia[2], Serigne Gueye[2], Micheli Knechtel[3], Philippe Michelon[3]

[1] Amadeus IT
{prénom.nom}@amadeus.com
[2] Université d'Avignon et des Pays de Vaucluse, LIA
{prénom.nom}@univ-avignon.fr
[3] Université d'Avignon et des Pays de Vaucluse, LMA
{prénom.nom}@univ-avignon.fr

Mots-clés : *Formulation linéaire, Relaxation Lagrangienne, Flux de passagers.*

1 Introduction

In 2013, 9.4 million flights have transported 842 million passengers and 13.4 million tons of freight and mail. Air traffic has increased considerably in the last years and it is expected that this trend will continue : an annual increase of 2.5% in the number of flights is expected until 2021 and a total increase of 50% is estimated for 2035, bringing to a total of 14.4 million flights[2, 1]. It is therefore essential that airlines companies, service companies (as Amadeus) and national and international regulatory agencies acquire planning and control tools. Nevertheless, traffic growth and increased number of passenger induce new problems of increasingly larger sizes.

In this paper, we propose to study one of these issues consisting in estimating the worldwide number of passengers by origin-destination pairs, on a monthly basis. If the number of passengers per flight may be estimated by statistical methods, they do not allow directly to deduce the number of travellers by origin-destination since several itineraries are generally available for a given pair OD, with the consequence that on a given flight, the passengers have different origins and destinations.

Estimating the number of passengers per O-D could allow to analyse the evolution of the demand for each O-D pair, estimate tourism flows entering or leaving a given city and anticipate the spread of infectious diseases such as Ebola or Zika.

2 Problem definition and model

Deducting an Origin-Destination matrix from partial data on segments doesn't constitute a new issue by itself. For example, one can cite the Bierlaire's works, which give a clear survey of the various existing approaches, [3, 4, 5]. It is however completely new in the field of aviation and there are two main reasons for this. The size of the problem in the airline industry is absolutely huge and vastly superior to the inherent problems in other areas, making the resolution by traditional methods unattractive. The second reason is related to obtaining data on a global scale, covering all airlines and airports (whose number exceeds three thousand, constituting more than ten million potential O-D pairs), which are unavailable for most of the companies, except services providers who work with all the airlines companies, such as Amadeus.

Specifically, our problem can be stated as follows :
Knowing the flow of passengers leaving from each airport, the flow of passengers arriving at the airports, an estimated number of passengers on each flight, lower bounds (limit below which

the flight is cancelled) and upper (capacity the plane) on the number of passengers that can be transported on the flight, the possible itineraries for each O-D pairs and the probability of using them (again estimated by statistical methods), find the number passenger for each O-D pairs.

Let now :

— a_i (resp. s_i) be the total number of passengers arriving at (resp. leaving) airport i,
— α^l_{od} be the proportion of passengers using leg l for going from o to d,
— $\hat{P}_l$ be an estimation of the number of passengers on leg l,
— $\underline{P_l}$ and $\overline{P_l}$ be the lower and upper bounds on the number of passengers on leg l,
— A (resp. L) be the total number of airports (resp. legs).

Let us define two sets of decision variables (although the second set of variables could be removed) :

— X_{od} : flow of passengers from o to d.
— P_l : number of passengers on leg l.

The problem can then be modelled as follows :

$$\begin{array}{lllr}
\min & \sum_{l=1}^{L} \beta_l (P_l - \hat{P}_l)^2 & & \\
\text{s.t.} & \sum_{d=1}^{A} X_{od} = s_o & \forall o \in \{1, ..., A\} & (1) \\
& \sum_{o=1}^{A} X_{od} = a_d & \forall d \in \{1, ..., A\} & (2) \\
& P_l = \sum_{(o,d) \in \{1,...,A\}^2} \alpha^l_{od} X_{od} & \forall l \in \{1, ..., L\} & (3) \\
& \underline{P_l} \leq P_l \leq \overline{P_l} & \forall l \in \{1, ..., L\} & (4) \\
& X_{od} \geq 0 & \forall (o,d) \in \{1, ..., A\}^2 &
\end{array}$$

Constraints (1) and (2) ensure that the number of passengers arriving to and leaving the airports are equal to the corresponding data. Constraints (3) compute the number of passengers per leg as a function of the flows of passengers, while constraints (4) and (5) indicate the domains of the decision variables. The objective function minimizes a weighted quadratic error with respect to the expected number of passengers per flight.

As a consequence, the whole problem consists in minimizing a convex quadratic (and separable) objective function subject to linear constraints. Such a problem does not present any particular theoretical difficulties. Nevertheless, worldwide instances include more than 3300 airports, leading to more than ten millions of O-D pairs! The challenge in solving this non-linear problem is thus to deal with its huge size.

We propose a Linear Programming based approach for solving the problem by substituting the quadratic error by absolute values. Furthermore, a closer look to the problem shows that constraints (3) actually links the X_{od} and the P_l variables. Thus, we also propose a Lagrangean Relaxation approach by associating a dual variable λ_l to each of these linking constraints and sending them in the objectif function.

Références

[1] Michel Bierlaire. The total demand scale : a new measure of quality for static and dynamic origin–destination trip tables. *Transportation Research Part B : Methodological*, 36(9) :837–850, 2002.

[2] Michel Bierlaire and Ph L Toint. Meuse : An origin-destination matrix estimator that exploits structure. *Transportation Research Part B : Methodological*, 29(1) :47–60, 1995.