Optimization model to design aviation networks
Greek PSO case study

Francisco Costa Pereira Leandro
francisco.leandro@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa – Portugal
November 2019

Abstract. Aviation has been facing a constant expansion since the first flight by an airplane was achieved, and nowadays the success of aviation, is closely linked with the success of local economies. On the other hand, the airline business is a challenging environment to thrive economically, due to the high costs and low profit margins involved. Hence, it is critical that airlines optimize their operation, in order to succeed in the long term, in this highly competitive business. Several models have been proposed in the literature to support the optimization of airline fleets, with the objective of minimizing the operational cost for the airline. On the other hand, the amount of literature dedicated to the optimization of the usage of airline's fleets dedicated to Public Service Obligation (PSO) routes is much sparser. Based on the research developed by Pita et al. (2013), with case studies applied to the PSO networks of the Azores and Norway, this model is adapted, the objective of which is not only to minimize the cost to the airline, but also to minimize the total social costs. Then, the model is applied to a new case study based on a PSO network for the Greek islands, with key differences from the previous case studies such as strong competition from the ferry boat service.

Keywords: Public Service Obligation; Greek Islands; Flight Scheduling and Fleet Assignment; Integer Linear Programming; Optimization

1. Introduction

Airports are key drivers of economic development for their respective catchment area([1], [2], [3]). This is even visible in a global context, in extreme cases such as Dubai, which is nowadays a thriving emirate, due to the strategy in investing in their airport and airline carrier to make it a global hub for airline transport [4].

Due to the close dependence between the performance of the aviation business and economies, it is very important to set conditions that will allow this business to thrive. Besides this, the airline business is known to operate in challenging conditions, which is explained by the high costs and low profit margins involved. Hence, it is crucial for an airline's long-term survival to use its resources in the most optimal way possible. There is extensive literature analyzing options to increase the operating margins of airlines (such as [5], [6], [7]).

Also, there is extensive literature addressing specifically the problem of optimizing the airlines' flight scheduling and fleet assignment (FSFA). The objective of such literature is to reduce the total operational cost for the airline. The result is usually the suppression of frequencies in less profitable routes and allocation of the resources to the most profitable ones.

However, there are other routes, whose main objective is not to maximize profits but to provide accessibility to remote areas, where there is not enough demand for profitable airline operation. Nonetheless, this operation is vital for local communities, and in these networks, the objective is not only to minimize costs, but also to maximize the quality of the service provided to passengers. In this research area, there is comparatively less published literature, with an opportunity for important research.

Integrated Flight Scheduling and Fleet Assignment Problem

Although the airlines that operate within PSO networks are financially rewarded for their service, it remains critical for them to operate as efficiently as possible. On the other hand, the entity responsible for subsidizing the PSO network is focused on maximizing the quality of service provided to the users, but also in minimizing the cost of subsidies that it must provide. Hence, solving the Integrated Flight Scheduling Fleet Assignment problem (IFSFA), is the suitable tool to optimize such networks.

The IFSFA model builds on the FSFA (which is only focused in cost reduction to the airline) by adding the minimization of the costs associated with the passenger, allowing for a more holistic view of the concept of an optimal network, not only reducing costs, but at the same time maximizing the quality of the service provided. This is very important because these networks are, by definition, sub-optimal, due to the fact that flights are being imposed on routes that do not have enough demand to justify such route.

This problem has already been explored previously and published in the literature ([8] and [9]) with positive results, reducing both financial costs for the airline and time costs for the passengers. In the present work, it will be applied to a different network with key differences, such as: the significant seasonality effect of demand (increasing in summer months due to tourism), or the significant competition from the boat service (an established transport in the region).

2. Two Greek case studies

This research applies an optimization model to two case studies located in Greece, within the Greek PSO network. Hence, for each case study, one "based" in Rhodes and another "based" in Thessaloniki, the airports which had PSO routes imposed connecting them to the "hub" airports were joined into two networks.

Each network is comprised of 8 airports, including the "hub", with 56 possible routes. These networks were chosen because, although they have the same amount of airports and are located relatively close to each other, they are different in terms of the amount of aircrafts employed, passengers transported, and frequencies imposed by the PSO, which will allow for a more comprehensive analysis of the Greek market.

The goal of these case studies is to reduce the total costs of the networks to the lowest possible values. The total cost to be considered is the sum of the following four components:

1. Aircraft direct operating costs;
2. Aircraft ground costs;
3. Passenger on board time costs;
4. Passenger ground connection time costs.

Rhodes Network

The first case study has Rhodes airport as its "hub". It can be considered the most simple network due to the smaller number of frequencies imposed, aircraft operating and the overall smaller costs involved, when compared to the second case study.

In this network there are 7 routes imposed by the PSO. The fleet that operates this network is composed by one Bombardier Dash 8 Q100 aircraft and two ATR 42 aircrafts. This results from an extensive analysis of the aircrafts operating in these routes, through flight tracking websites, leading to the conclusion that this was the most accurate representation of the real situation.

Costs

Regarding the aircraft operating costs, from data provided by the HCAA from last summer, the number of movements and aircraft type associated with each origin-destination (O/D) pair was noted. Using this data multiplied by the duration of each flight, the total aircraft operating cost was estimated to be 53 771€, for a total of 50 flights, resulting in an average cost of 1 075€ per flight.

Regarding the aircraft ground costs, they are considered to be the parking fees in the airport. It is common in the airline industry for airlines not to pay parking fees if the aircraft is on the ground for at least less than 2 hours [10]. Since these airlines schedule flights to avoid having aircrafts on the ground for more than 2 hours, it was considered that for the current network, aircraft ground costs are zero.

Regarding the Passenger time costs, the total travel time for each O/D pair was compiled. These values were compiled through an extensive search from online travel websites, and for each O/D pair, the travel times of at least one full week were verified, and the shortest value was considered. This was done in order to allow for a fair comparison, due to the fact that there are days of the week which allow for better connections than others. Some routes have direct flights, whereas others require long connections, explaining the broad range of values for the travel time.

With this information, for each O/D pair the passenger time costs were estimated using the following expression:

$$PTC = Ct \times TT \times PN \qquad (1)$$

Where Ct is the cost of time for the passengers, TT the travel time and PN the number of passengers in that route. The total passenger time costs were then estimated by summing all the passenger time costs for each O/D pair associated with that network. The total passenger time costs were estimated to be 6 470€ for a total of 375 passengers, resulting in an average travel time of 1 hour and 44 minutes per passenger.

Summing all these components, the total cost for the Rhodes network was estimated to be 60 241€.

Thessaloniki network

This network is also comprised of 7 routes. These are operated by 2 dash 8 Q100, 2 dash 8 Q400 and 1 ATR 42 aircrafts. The composition of the fleet in this network was obtained through the same process as in the previous network.

Costs

The same method was used to calculate the aircraft operating costs of this network. The total aircraft operating costs were estimated to be 134 424€, for a total of 62 flights, resulting in an average cost of 2 168€ per flight (which is more than twice the average operating cost for the Rhodes network).

For the same reasons as explained for the Rhodes network, the aircraft ground costs for the Thessaloniki network were assumed to be 0€.

Regarding the Passenger time costs, the total travel time for each O/D pair was compiled using the same method as in the previous network. Using equation (1), the total passenger time costs for the Thessaloniki network were estimated to be 27 637€ for a total of 1357 passengers (three times more passengers than in the Rhodes network), resulting in an average travel time of 2 hours and 1 minute per passenger.

Summing all the components above, the total cost for the Thessaloniki network was estimated to be 162 061€.

3. Related Literature Review

Demand Prediction

Grosche et al. [10] proposed two possible gravity models, underlining the known fact that there is considerable unreliability with this task. The model is applicable to new markets, with the advantage of not relying on inputs that are not yet available to airlines before starting to operate the route. Instead, the model uses mainly geo-economic variables as independent factors.

Wadud [11] published a paper analyzing demand prediction in areas where data to define explanatory variables is not available. This research highlights the challenges and uncertainties associated with the demand prediction task and performs it with limited data available, also taking into account the competition by road travel, which can be important in relatively small countries.

Kluge et al. [12] perform an analysis applied specifically to the European market. The paper focuses on determining the relation between passenger air travel demand and factors such as the GDP, the urbanization level, the geographical location and the degree of education, proving that the first, third and fourth indicators were statistically significant.

An interesting niche of this research subject is the demand prediction for markets with very strong touristic activity, due to the differences that these carry with them (such as seasonality, less importance of GDP when compared to more traditional business markets, very high ratios of tourist to inhabitant, etc). This type of markets have already been analyzed since at least 2002, when Devoto et al. [13] published their research focused on determining how demand could be predicted in these touristic markets, specifically using tourism variables (e.g. resident population, number of tourist beds, per capita beds and tourist arrivals). More recently, Erjongmanee & Kongsamutr [14] published a research focusing on demand forecasting in Thailand, taking into account the effect of tourism, with significant results as predicted.

Fleet Optimization

Lohatepanont & Barnhart [15] and Sherali et al. [16] are two widely recognized publications that assess the problem of flight scheduling and fleet assignment with the sole purpose of maximizing profit for airlines. These papers obtained interesting improvements in their case studies, with several publications building on this objective. One of these, published more recently, is Jamili [17] , which has the same objective, although exploring different methodologies to achieve it.

The main reference for this work came from Pita et al. ([8] and [9]), where a model was built that, instead of focusing on maximizing economic results for airlines, adds the objective of the maximization of the quality of the service provided to passengers. The models are applied to case studies in the PSO networks established in the Azores and in Norway, respectively. The second paper, builds on the first one, taking also into account the expenses and revenues of airport owners, associated with these routes. Both case studies obtained very interesting results, reducing costs in all the areas considered, and with impressive computational times required to reach the optimum solution.

Continuing the previously mentioned research, Antunes et al. [18] focused on analyzing in depth the network of the Azores operated by SATA, working closely with the airline. This allowed real data to be used as much as possible, reducing the amount of assumptions. With the objective of analyzing the maximum reduction of operating costs that SATA could have achieved by optimizing the network and changing its route structure. This was done while satisfying the same passenger demand as in 2012, taking into account the implications of possible changes to the level of service offered. The paper proposed new shapes for the PSO imposed network, and quantified the improvements that could be obtained, with real data from a year in the past. The research concluded that the variable operating costs could be reduced significantly, which would save the government of the Azores a significant amount of funds in subsidies.

Iliopoulou et al. [19] was analyzed due to the similarities it has with the present research. This paper proposes a sea-plane network in the Greek islands, which would compete against the locally well-established boat network. The objective is to minimize the travel cost, the size of the fleet and the unsatisfied demand between successive island ports, by proposing a new network, instead of optimizing an existent route.

Ma et al. (2017) [20] addresses arguably one of the most discussed problems currently, which is the need for the reduction of carbon dioxide emissions. It develops an optimization model whose objective is simultaneously maximize profit and minimize emissions associated with operating the flights. This model is applied to case studies from Asian airlines, with interesting results, namely that the optimal point obtained mathematically proved to be unreachable in real life. Besides this, the point achievable in reality that was closest to optimality had significant improvements over the current situation, and it was concluded that small reductions in profits lead to significant reductions in emissions.

4. Construction of the optimization model

The integer linear programming model will be discussed next.

Decision variables

The variables whose values will be optimized when running the model, the decision variables, are:

1. y(a,t,r): number of aircrafts of type r that are on the ground in airport a, from time t to t+1;
2. x(f,t,r): number of aircrafts of type r that fly route f, departing at time t and arriving at time [t+t_F(f)], this variable is imposed to be binary;
3. u_D(f,t): number of passengers assigned to route f, taking off at t and landing at [t+t_F(f)];
4. u_1(f,a,t,wt): number of passengers assigned to the one stop itinerary which contains route f, and then continues to final destination a. Initial departure time is t, waiting time on the ground for connection is wt. Hence, the time of final arrival is given by [t+t_F(f)+wt+t_A(a_F(f),a)].
5. u_2(f_1,f_2,t,wt_1,wt_2): number of passengers assigned to the two stop itinerary which contains f_1 as the first flight and f_2 as the third flight, and has a flight joining the two airports as the second flight. Initial departure time is t and the waiting times on the ground are respectively wt_1 and wt_2. Hence, the time of final arrival is given by: [t+t_F(f_1)+wt_1+t_A(a_F(f_1),d_F(f_2))+wt_2+t_F(f_2)];
6. g_c(a,t,r): equal to 1 if aircraft r has been on the ground for more than 2 hours, starting at time t;

It should be noted that this is not the most straightforward formulation, but with 56 routes, 8 airports and 33 time periods, initial simpler formulations would quickly have too many indexes for a system with 8 Gigabytes (GB) of random access memory (RAM) to run out of memory (e.g. a reduction in the 10^6 order of magnitude was obtained in the amount of entries of u_2).

Objective function

It is critical that the objective function is properly defined, and correctly reflects the physical reality of the problem. The objective function was defined as the minimization of the sum of seven components (O_1 to O_7). It should be noted that the equations of the objective function were placed at the end of the document, in the appendix, in order to improve their readability, due to space limitations.

The first component (2) reflects the direct costs for the airline, resulting from the operation of the flights.

The second component (3) accounts for the costs to the airline, of having an aircraft of type r, parked on the ground in airport a, at time t (when it exceeds 2 hours).

The third, fourth and fifth components [(4), (5), (6)] account for the social cost to the passengers, quantified by the cost of time for them, for direct, one stop and two stops itineraries, respectively. The sixth and seventh components [(7), (8)] account for the social cost to the passengers, of having to wait between two flights, on the ground, respectively for one and two stop itineraries. Hence, it is obtained by summing the product of all the time on the ground in an airport, with the number of passengers and the cost of time on the ground for passengers.

Constraints

For the same reason as the objective function's equations, those which define the constraints are also presented in the appendix.

The first constraint (9) ensures that the sum of aircrafts on the ground and in the air, at any time period, is equal to the available number of aircrafts of that type.

The second constraint (10) imposes continuity in each node. It imposes that the sum of the number of aircrafts arriving into an airport and aircrafts already parked there, is equal to the sum of aircrafts departing that airport and aircrafts that will stay parked there.

Constraint (11) imposes that there are never more passengers assigned to a flight than the maximum allowed number of passengers to that flight. This is achieved by specifying that for all aircraft types, the sum of all passengers in direct or connecting flights is smaller or equal to the number of available seats.

Constraint (12) ensures that the demand is satisfied, i.e. that all the passengers that must travel from one airport to another, will either be assigned to a direct, one-stop or two-stop itinerary.

Constraints (13) and (14) impose that, respectively, the minimum number of flights and seats as imposed by the PSO, between any two airports is fulfilled.

Constraints (15), (16), (17), (18) and (19) impose that, respectively, the number of aircrafts on the ground, in the air, passengers carried in direct, one-stop and two-stops itineraries are all positive integers.

Constraints (20) and (21) impose that the fleet starts and ends the day at the hub. Constraints (22) and (23) allow the model to only consider aircraft ground fees if an aircraft stays on the ground for more than 2 hours. Constraint (24) imposes that there are not two different flights operating on the same route, with an interval smaller than 3 hours. This had to be imposed because one solution fulfilled all the frequencies imposed by the PSO for a specific route with very small intervals, which is unreal.

Besides the above-mentioned constraints, which are necessary for the correct specification of the problem, other "virtual" constraints were added.

The goal was reducing the computation time to reasonable values and these constraints were added based on the concept of "helping" the model in narrowing down the range of possible solutions only to the reasonable ones. This removes from the scope of analysis of the software unreasonable solutions, such as placing passengers in itineraries which end the day in the same airport as that of departure.

This specification of additional "virtual constraints" must be carried out carefully, under the risk of removing the actual optimal solution from the range of possible solutions to be analyzed by the model.

Some examples of these constraints which were attempted, some with and some without success are:

1. Whenever one aircraft is departing the "hub" airport, all the fleet is departing the "hub" airport at that time. This potentializes the hub effect, and increases the possibility of connections in the hub, requiring less flights overall;
2. Imposing that in any moment in time there is a maximum of one aircraft operating in each route;
3. Specifying a maximum of one flight for the whole time of the analysis, for all the routes that have no minimum amount of flights assigned by the PSO network, or have low demand;
4. Impose that connecting itineraries which imply a total flown distance longer than 150% of the direct distance between O/D do not have passengers placed there.

After specifying the optimization model, an illustrative example was defined, in order to verify its correct specification, with positive results.

5. Predictive model

The demand for the O/D pairs was estimated through multiple variable linear regression analysis, since this is a method commonly accepted in the literature concerning this area. With the objective of using published literature as a guideline, a literature review was carried out through published papers which analyze the problem of demand prediction, in order to choose the most suitable explanatory variables to the case study. A summary of the most common explanatory variables found in these publications was defined, and is presented below:

1. GDP (either summing or multiplying both origin and destination, either total or per capita);
2. Population (either summing or multiplying both origin and destination);
3. Importance of tourism (either by number of tourist arrivals, hotel beds or per capita beds);
4. Cost of ticket (either absolute, or compared to its competition (e.g. rail, car, boat…));
5. Travel time;
6. Distance between airports.

Since this case study has some particularities, other variables that could describe them were considered, and later its significance was assessed through the multiple variable linear regression analysis, such as:

1. Existence of significant cruise ship terminals in the islands, since it is expected that the embarking and disembarking of cruise ship passengers will increase the demand of passengers for airliners;
2. Competition of the ferry boats, since this is a well-established and very popular mean of transportation within the Greek archipelago.
3. Effect of population ageing. The hypothesis that areas with a higher share of retired population would have proportionally less travelling will be tested. Since the average age of the population of some islands is above the Greek average, this was considered.

The first step in the multiple variable linear regression analysis was the data collection, namely the values of the above explanatory variables, for each of the 240 O/D pairs (for the predictive model, both networks were joined into one). This data was collected through several sources, and is applicable to the month of August 2018.
Once this information was collected, the regression was carried out using IBM's SPSS software package, through a Poisson regression.
This type of regression was chosen due to its greater suitability to these types of data sets, with very different values of the dependent variable (demand), which depend on explanatory variables by a power different than one.
The regression went through several specification tests, in order to reach the most reasonable model possible, such as:

1. Checking for overdispersion of data, through the Lagrange Test, in order to validate either the Poisson regression or the negative binomial regression as the best option;

2. Verification of the statistical significance of the parameters, through the Wald test and p-values;
3. Analysis of the predictive capacity, through the Omnibus test;
4. Comparison between models with different specifications, in order to choose the most suitable one.

The key performance indicators on a small group of the best performing models was compiled into Table 1, in order to choose the final model to be used. Models 2 and 5 were those with the most promising indicators

After calculating this demand, the actual demand was compared with the predicted demand for the 66 routes. It was concluded that, although model 5's indicators suggested better performance, this model is strongly overestimating the demand for the smaller markets, and slightly overestimating for the bigger markets. Since the model is applicable to PSO routes, which are characterized by low demand, the decision that model 2 was the most suitable choice was made, due to the fact that the predicted values are closer to the actual values.

Table 1: KPI's of the different predictive model candidates

Model	M1		M2		M3		M4		M5	
offset included?	No		No		Yes		No		Yes	
	Beta	Std Dev.	Beta	Std Dev.	Beta	Std Dev.	Beta	Std Dev.	Beta	Std Dev.
Log Population Product	0.56***	0.12	0.30**	0.14	0.30**	0.096	0.47***	0.11	0.19*	0.09
Distance	0.01*	0.01	0.01***	0.01	0.01***	0.002	0.01**	0.01	0.01***	0.01
Frequency of flights	0.09***	0.02	0.10***	0.02	0.10***	0.017	0.07***	0.02	0.07***	0.02
Cost ticket air	-0.01*	0.01	---------	----------	---------	----------	-0.01*	0.01	---------	----------
Big market	---------	----------	0.10***	0.29	---------	----------	---------	----------	---------	----------
Travel time	---------	----------	---------	----------	---------	----------	-0.46*	0.21	-0.55**	0.20
Goodness of fit:										
AIC	8710.9		7414.4		7412.4		7749.1		6339.7	
log-likelihood	-4350.5		-3702.2		-3702.2		-3868.6		-3164.8	
Deviance	8314.1		7017.5		7017.5		7350.3		5942.8	

6. Results and discussion

After running the optimization through the Fico Xpress software package, satisfactory results were obtained in both networks. They will be presented and discussed in the following paragraphs, one network at a time. As a reference, the calculation was performed in a Windows 10 Pro operating system, running in a computer with an Intel(R) Core(TM) i7-3770K CPU @ 3.50 GHz, and 8 GB of RAM memory

Rhodes Network

This network is considerably smaller than the network "based" in Thessaloniki (total costs are 42% of the value of the Thessaloniki network), and was less demanding to solve, with respect to computational effort. At the beginning, with all the inputs for the model defined, "virtual constraints" were added, with the objective of accelerating the convergence towards the optimal solution. Examples of these virtual constraints are restricting the number of flights for each route. Another virtual constraint that was attempted, was imposing that for every period of time, either no aircraft would depart the hub, or all the fleet would depart the hub, in an effort to promote the "hub effect", and increase the number of passengers whose itinerary would be satisfied by connecting flights, reducing the total amount of flights.

Unfortunately, although this technique reduced the computational times required for the solutions to be determined by the software significantly, after carefully analyzing the solutions, it was decided that these "virtual constraints" were not valid, because they were removing from the range of possible solutions, solutions with lower total costs, besides following all the real constraints. Hence, the decision to remove these virtual constraints was taken, with the purpose of achieving the real optimal solution, at the expense of longer computational solving periods.

The optimal solution for the Rhodes network was obtained after 14 hours and 41 minutes, with an optimality gap of 8.58%. A schematic representation of the flights in the optimized network, including passengers on board was included in the appendix. The costs associated with these solution, and the comparison with the current network's costs are:

- **Flight operating costs**: 48 815€ for a total of 46 flights, compared to 53 771€ for the current network, which requires 50 flights, a reduction of 9.2% in cost;

- **Aircraft ground costs**: 75€ compared with 0€ for the current network. This increase is considered negligible when compared to the other reductions obtained by the model.
- **Passenger time costs**: 3 995€ for the time passengers spent on board the aircraft, and 860€ for the time passengers spent waiting on the ground for a connecting flight, in a total of 4 855€ resulting in an average travel time of 1 hour and 17 minutes per passenger, compared with 6 470€ for the current network and an average travel time of 1 hour and 44 minutes per passenger. This means a reduction of 24.9% in cost;
- **Total cost of the network**: 53 745€ compared with 60 241€ for the current network. This means a reduction of 10.7% in the total cost of the network, with reductions in all the parameters, except a negligible increase from 0 to 75€ in the aircraft ground costs, fulfilling the objective of not only reducing the financial costs associated with the network, but also improving the quality of service provided to the passengers, through reduction of the door to door travel time.

Thessaloniki Network

This network is, as already mentioned, significantly larger than the one discussed above, hence, it required longer computational times to reach solutions. After realizing that applying the virtual constraints was actually excluding optimal and valid solutions from the solution domain, it was decided to run this optimization right from the start without applying the virtual constraints which were excluded from the previous network, with the objective of guaranteeing that the software would consider every valid solution, at the expense of longer computational times. The optimal solution was found after 5 hours and 30 minutes, with an optimality gap of 11.26%. The model was then left running for another 15 hours, without any improvement. As with the previous network, the schematic representation of the flights in the optimized network was included in the appendix.

The costs associated with this solution, and the comparison with the current network's costs are:

- **Flight operating costs**: 117 978€ for a total of 54 flights, compared to 134 424€ for the current network, which requires 62 flights, a reduction of 12.2% in cost;
- **Aircraft ground costs**: 225€ compared with 0€ for the current network. This increase is considered negligible when compared to the other reductions obtained by the model.
- **Passenger time costs**: 21 900€ for the time passengers spent on board the aircraft, and 3 565€ for the time passengers spent waiting on the ground for a connecting flight, in a total of 25 465€ resulting in an average travel time of 1 hour and 52 minutes per passenger, compared with 27 637€ for the current network and an average travel time of 2 hours and 02 minutes per passenger. This means a decrease of 7.86% in cost;
- **Total cost of the network**: 143 668€ compared with 162 061€ for the current network. This means a reduction of 11.3% in the total cost, having reduced once again both the direct financial costs to the airlines, as well as the time costs for the passengers.

Exploring scenarios

As mentioned before, there were two demand model candidates considered statistically viable, but model 2 was used in the optimization. In order to perform a sensitivity analysis, the optimization was also performed for the demand resulting from model 5. The results were positive, with improvements being obtained in all costs, except for the direct operating costs in the Thessaloniki network. The reason for this increase is believed to be the fact that there was an increase of 37% in demand, and the current network direct operating costs were kept constant, which is not likely to be a fair comparison.

7. Conclusions, Limitations and Future

Conclusions

The present work adapted a published optimization model ([8], [9]), in order to apply it to two case studies situated in the Greek PSO network. It demonstrated that the network can be improved, not only from a financial point of view, but also regarding passenger level of service.

One of the main strengths of this work is combining both the development of a predictive model and a flight scheduling and fleet optimization model.
The dataset presented challenges for the development of the Generalized linear model, such as significant overdispersion of data, which led to the rejected attempt to use a Negative binomial regression, followed by a Poisson regression with a Pearson chi-square scale parameter method. This resulted in the inability to use as many explanatory variables as initially intended, due to their statistical non-significance. Nevertheless, two demand models complied with the desired level of performance in the evaluated KPI's, and were tested in the optimization model.
The optimization resulted in significant improvements, in the order of 10% in both networks, while following all the constraints specified, in order to properly characterize the particularities of the Greek PSO market.

Limitations

This research holds some limitations, which should be acknowledged by the readers. Firstly, regarding the results of the predictive model, its inherent uncertainty should not be disregarded. This is partly explained by the lack of data available and by the low demand values for these O/D pairs and is demonstrated by the significant difference in the prediction between the two final candidates for demand model.

Moreover, regarding the optimization model, although 48 hours is an acceptable duration for running this computation, required twice a year, its convergence towards optimality is limited. This is demonstrated by the difficulty for the model to improve solutions. Also, the optimality gap of the calculations was around 8% which is not ideal.

Future Work

From the point of view of the predictive model, more attention into attempting to obtain significance from variables related to the importance of tourism and of the competition of the boat service is also an important opportunity.

Regarding the optimization model, one interesting opportunity, is to develop a similar research, but optimizing the network for the winter months, and using demand data from the same period for the regression. Then, an in depth comparison between both networks should bring interesting data, in a market with such a strong effect from seasonality, mostly due to tourism.

8. References

[1] ACI Europe, "The Social and Economic Impact of Airports," York Aviation, Geneva, 2004.

[2] Halpern, N. and Bråthen, S., "Impact of airports on regional accessibility and social development," *Journal of Transport Geography,* pp. 1145-1154, 2011.

[3] Lieshout, R., "Measuring the size of an airport's catchment area," *Jornal of Transport Geography,* vol. 25, pp. 27-34, 2012.

[4] Lohmann, G., Albers, S., Koch, B., Pavlovich, K., "From hub to tourist destination – An explorative study of Singapore and Dubai's aviation-based transformation," *Journal of Air Transport Management,* vol. 15, pp. 205-211, 2009.

[5] Zou, L., Chen, X., "The effect of code-sharing alliances on airline profitability," *Journal of Air Transport Management,* vol. 58, pp. 50-57, 2017.

[6] Raynes, C., Tsui, K.W.H., "Review of Airline-within-Airline strategy: Case studies of the Singapore Airlines Group and Qantas Group," *Case Studies on Transport Policy,* vol. 7, pp. 150-165, 2019.

[7] Merkert, R., Swidan, H., "Flying with(out) a safety net: Financial hedging in the airline industry," *Transportation Research Part E,* vol. 127, 2019.

[8] Pita, J.P., Antunes, A.P., Barnhart, C., Menezes, A.G., "Setting public service obligations in low-demand air transportation networks: Application to the Azores," *Transportation Research Part A,* vol. 54, pp. 35-48, 2013.

[9] Pita, J.P, Adler, N., Antunes, A.P., "Socially-oriented flight scheduling and fleet assignment model with an application to Norway," *Transportation Research Part B,* vol. 61, pp. 17-32, 2014.

[10] Grosche, T., Rothlauf, F., Heinzl, A., "Gravity models for airline passenger volume estimation," *Journal of Air Transport Management,* vol. 13, pp. 175-183, 2007.

[11] Wadud, Z., "Modeling and Forecasting Passenger Demand for a New Domestic Airport with Limited Data," *Transportation Research Record,* pp. 59-68, 2011.

[12] Kluge, U., Cook, A.J., Paul, A., Cristobal, S., "Factors influencing European passenger demand for air transport," in *Air Transport Research Society World Conference*, 2017.

[13] Devoto, R., Farci, C., Liliu, F., "Analysis and forecast of air transport demand in Sardinia's airports as a function of tourism variables," *Urban Transport VIII,* vol. 60, 2002.

[14] Erjongmanee, S., Kongsamutr, N., "Air Passenger Estimation Using Gravity Model and Learning Approaches: Case Study of Thailand," in *5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, Thailand, 2018.

[15] Lohatepanont, M., Barnhart, C., "Airline Schedule Planning: Integrated Models and Algorithms for Schedule Design and Fleet Assignment," *TRANSPORTATION SCIENCE,* vol. 38, pp. 19-32, 2004.

[16] Sherali, H.D., Bae, K.H., Haouari, M., "Integrated Airline Schedule Design and Fleet Assignment: Polyhedral Analysis and Benders' Decomposition Approach," *INFORMS Journal on Computing,* vol. 22, pp. 500-513, 2010.

[17] Jamili, A., "A robust mathematical model and heuristic algorithms for integrated aircraft routing and scheduling, with consideration of fleet assignment problem," *Journal of Air Transport Management,* vol. 58, pp. 21-30, 2016.

[18] Antunes, A.P., Santos, M.G., Pita, J.P., Menezes, A.G., "Study on the evolution of the air transport network of the Azores," *Transportation Research Part A,* vol. 118, pp. 837-851, 2018.

[19] Iliopoulou, C., Kepaptsoglou, K., Karlaftis, M.G., "Route planning for a seaplane service: The case of the Greek Islands," *Computers&OperationsResearch,* vol. 59, pp. 66-77, 2015.

[20] Ma, Q., Song, H., Zhu, W., "Low-carbon airline fleet assignment: A compromise approach," *Journal of Air Transport Management,* vol. 68, pp. 86-102, 2017.

Appendix

$$\mathrm{O}_1 = \sum_{\mathrm{f} \in \mathrm{FL}, \mathrm{t} \in \mathrm{T}, \mathrm{r} \in \mathrm{R}} \mathrm{c}_F(\mathrm{r}).\, t_F(\mathrm{f}).\, \mathrm{x}(\mathrm{f}, \mathrm{t}, \mathrm{r}) \tag{2}$$

$$\mathrm{O}_2 = \sum_{\mathrm{a} \in \mathrm{A}, \mathrm{t} \in \mathrm{T}, \mathrm{r} \in \mathrm{R}} \mathrm{c}_S(\mathrm{a}, \mathrm{r}).\, \mathrm{g}_c(\mathrm{a}, \mathrm{t}, \mathrm{r}) \tag{3}$$

$$\mathrm{O}_3 = \sum_{\mathrm{f} \in \mathrm{FL}, \mathrm{t} \in \mathrm{T}} \mathrm{c}_B.\, \mathrm{t}_F(\mathrm{f}).\, \mathrm{u}_D(\mathrm{f}, \mathrm{t}) \tag{5}$$

$$\mathrm{O}_4 = \sum_{\substack{\mathrm{f} \in \mathrm{FL}, \mathrm{a} \in \mathrm{A} \\ \mathrm{t} \in \mathrm{T}, \mathrm{wt} \in \mathrm{WT}}} \mathrm{c}_B.\, [\mathrm{t}_F(\mathrm{f}) + \mathrm{t}_A(\mathrm{a}_F(\mathrm{f}), \mathrm{a})].\, \mathrm{u}_1(\mathrm{f}, \mathrm{a}, \mathrm{t}, \mathrm{wt}) \tag{4}$$

$$\mathrm{O}_5 = \sum_{\substack{\{\mathrm{f}_1, \mathrm{f}_2\} \in \mathrm{FL}, \mathrm{t} \in \mathrm{T}, \\ \{wt_1, wt_2\} \in \mathrm{WT}}} c_B.\, \left[t_F(f_1) + t_A\big(a_F(f_1), d_F(f_2)\big) + t_F(f_2)\right].\, u_2(f_1, f_2, t, wt_1, wt_2) \tag{6}$$

$$\mathrm{O}_6 = \sum_{\substack{\mathrm{f} \in \mathrm{FL}, \mathrm{a} \in \mathrm{A} \\ \mathrm{t} \in \mathrm{T}, \mathrm{wt} \in \mathrm{WT}}} \mathrm{c}_W.\, \mathrm{wt}.\, \mathrm{u}_1(\mathrm{f}, \mathrm{a}, \mathrm{t}, \mathrm{wt}) \tag{7}$$

$$\mathrm{O}_7 = \sum_{\substack{\{\mathrm{f}_1, \mathrm{f}_2\} \in \mathrm{FL}, \mathrm{t} \in \mathrm{T}, \\ \{wt_1, wt_2\} \in \mathrm{WT}}} \mathrm{c}_W.\, (\mathrm{wt}_1 + \mathrm{wt}_2).\, \mathrm{u}_2(\mathrm{f}_1, \mathrm{f}_2, \mathrm{t}, \mathrm{wt}_1, \mathrm{wt}_2) \tag{10}$$

$$\sum_{a \in A} y(a, t, r) + \sum_{\substack{f \in FL, t_1 \in T | \\ t_1 \leq t < (t_1 + t_F(f))}} x(f, t_1, r) = z(r), \qquad \forall\, t \in T, r \in R \tag{11}$$

$$y(a, t-1, r) + \sum_{\substack{f \in FL | \\ a_F(f) = a \wedge t > t_F(f)}} x(f, t - t_F(f), r) = y(a, t, r) + \sum_{\substack{f \in FL | \\ d_F(f) = a}} x(f, t, r) \quad , \forall a \in A, t \in T \setminus \{1\}, r \in R \tag{9}$$

$$\begin{aligned} \sum_{r \in R} l(f).\, s(r).\, x(f, t, r) \geq u_D(f, t) + \sum_{a \in A, wt \in WT} u_1(f, a, t, wt) + \sum_{\substack{f_1 \in FL,\ a \in A, t_1 \in T,\ wt \in WT | \\ a_F(f_1) = d_F(f) \wedge a_F(f) = a \wedge (t_1 + t_F(f_1) + wt) = t}} u_1(f_1, a, t_1, wt) \\ + \sum_{f_1 \in FL, \{wt_1; wt_2\} \in WT} u_2(f, f_1, t, wt_1, wt_2) + \sum_{\substack{\{f_1; f_2\} \in FL,\ t1 \in T,\ \{wt_1; wt_2\} \in WT | \\ a_F(f_1) = d_F(f) \wedge d_F(f_2) = a_F(f) \wedge (t_1 + t_F(f_1) + wt_1) = t}} u_2(f_1, f_2, t_1, wt_1, wt_2) \\ + \sum_{\substack{f_1 \in FL,\ t_1 \in T,\ \{wt_1; wt_2\} \in WT | \\ (t_1 + t_F(f_1) + wt_1 + t_A(a_F(f_1), d_F(f)) + wt_2) = t}} u_2(f_1, f, t_1, wt_1, wt_2), \quad \forall\, f \in FL, t \in T \end{aligned} \tag{8}$$

$$\sum_{t\in T} u_D(f,t) + \sum_{\substack{f_1\in FL, t\in T, wt\in WT, a\in A| \\ d_F(f)=f_1 \wedge a_F(f)=a}} u_1(f_1,a,t,wt) + \sum_{\substack{\{f_1;f_2\}\in FL, t\in T, \{wt_1;wt_2\}\in WT| \\ d_F(f)=d_F(f_1) \wedge a_F(f)=a_F(f_2)}} u_2(f_1,f_2,t,wt_1,wt_2) = q(f), \quad \forall f \in FL \tag{13}$$

$$\sum_{t\in T,\ r\in R} x(f,t,r) \geq x_{min}(f) \quad ,\forall f \in FL \tag{14}$$

$$\sum_{t\in T,\ r\in R} s(r).x(f,t,r) \geq s_{min}(f) \quad ,\forall f \in FL \tag{12}$$

$$y(a,t,r) \in \mathbb{Z} \quad ,\forall\, a \in A, t \in T, r \in R \tag{15}$$

$$x(f,t,r) \in \mathbb{Z} \quad ,\forall\, f \in FL, t \in T, r \in R \tag{16}$$

$$u_D(f,t) \in \mathbb{Z} \quad ,\forall\, f \in FL, t \in T \tag{18}$$

$$u_1(f,a,t,wt) \in \mathbb{Z} \quad ,\forall\, f \in FL, a \in A, t \in T, wt \in WT \tag{17}$$

$$u_2(f_1,f_2,t,wt_1,wt_2) \in \mathbb{Z} \quad ,\forall\, \{f_1;f_2\} \in FL, t \in T, \{wt_1;wt_2\} \in WT \tag{20}$$

$$\sum_{f\in FL|d_F(f)=8} x(f,1,r) + y(8,1,r) = z(r) \quad ,\forall\, r \in R \tag{19}$$

$$y(8,33,r) = z(r) \quad ,\forall\, r \in R \tag{21}$$

$$g_C(a,t,r) \in \{0,1\} \quad ,\forall\, a \in A, t \in T, r \in R \tag{22}$$

$$g_C(a,t,r) \geq y(a,t,r) + y(a,t+1,r) + y(a,t+2,r) + y(a,t+3,r) - 3.5 \quad ,\forall\, \mathrm{a} \in A, r \in R, t \in T \setminus \{31,32,33\} \tag{23}$$

$$\sum_{r\in R} [x(f,t,r) + x(f,t+1,r) + x(f,t+2,r) + x(f,t+3,r) + x(f,t+4,r) + x(f,t+5,r)] \leq 1, \quad \forall\, f \in FL, t \in T \setminus \{29,30,31,32,33\} \tag{24}$$