Robust Network Design for Air Transport under PSO

Alice da Silva Costa Lourenço
alice.silva.costa.lourenco@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal

June 2022

Abstract

Air transport is essential for social and economic development and cohesion of regions, particularly in geographically remote areas, like islands, though some routes are unprofitable due to low passenger demand. Public Service Obligations (PSO) aim to preserve those routes, assuring minimum transportation services to communities. The present work aims to develop robust tools to support the decision-making process led by the responsible authorities regarding air transport subsidized networks. Thereby, a Passenger Demand Prediction model and a robust Route and Flight Schedule Optimization program were developed. Given the computational complexity of this optimization problem, two relaxation approaches were also introduced. The proposed approach was applied to the Rhodes (Greece) based PSO network case study. First, the demand prediction results indicate that the demand in this network is affected by the population of the islands, touristic attractiveness, seasonality, and sea transport competition. It was also shown that the Covid-19 pandemic affected the impact of some of these explanatory variables. Second, it was confirmed that making only use of the expected demand to optimize the network yields a less robust solution, incapable of serving a plausibly optimistic scenario. This study contributes to the literature by accounting for different passenger demand scenarios simultaneously upon the optimization which is shown to be vital, since a deterministic approach is unable to adapt to uncertainty inherent to demand prediction.
Keywords: Public Service Obligation; Greek Islands; Robust Optimization; Air Transport Network Design; Passenger Demand Prediction

1. Introduction

One of the main concerns of air transport liberalization is that airlines invest more on already profitable routes and discard routes with insufficient passenger demand. This concentration leads to the underdevelopment of remote areas. The European Union (EU) defined Public Service Obligations (PSOs), aiming to preserve those vital yet unprofitable routes, assuring minimum transportation services to remote communities, so that residents have access to a transport system that confers connectivity to main urban areas. This subsidy scheme should keep a fair competition basis in the market and ensure that the financial support given does not turn out as a burden for taxpayers.

PSOs may be imposed by Member States on domestic and intra-EU routes. Authorities release an open public tender, specifying the PSO standards, to which any airline can apply. The selection process accounts for the adequacy of the proposed service (including the prices and conditions which can be offered to users) and for the subsidy amount required (if any) by the airline from the Member State. This work focuses on the definition of the minimum service standards under which a pre-defined subsidized network should be operated by optimizing the network's route and flight schedule. It is essential to acknowledge the great uncertainty inherent to passenger behavior and exogenous factors and, with this in mind, apply robust decision-making methods to design the air transport network.

The different objectives of all stakeholders (passengers, airlines, and government) should be considered. Within a decision support framework similar to the one schematized in Figure 1, the present work presents two main contributions: i) the development of passenger demand prediction models and ii) a robust Route and Flight Schedule Optimization program, that takes several passenger demand scenarios into account simultaneously.

2. Literature Review

The literature on both passenger demand prediction and air network design and operational optimization was explored. However, only subsidized networks optimization state-of-the-art is hereafter described.

An Integrated Flight Schedule and Fleet Assign-

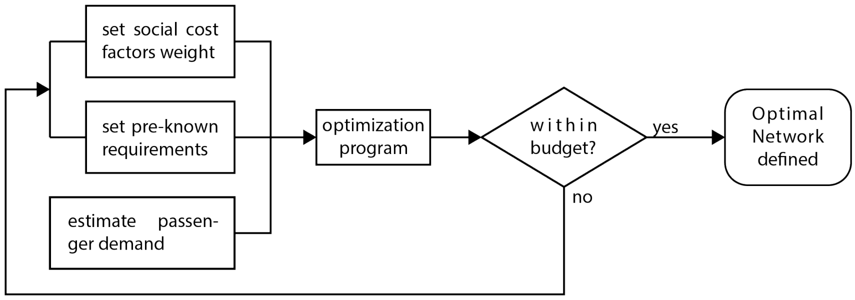


Figure 1: Example for a decision support framework.

ment (IFSFA) model is developed in [8], minimizing total costs accounting for the interests of passengers, airlines, and government. [7] builds up on the previous paper by including airport costs and applying it to a PSO network in Norway. The designated Socially-oriented FSFA (SFSFA) model is applied to a single operational day divided into 15 or 30 minuntes time steps and considers itineraries of up to two stops. Xpress (Fico, 2011) commercial software was used and the computation time required to reach optimal solutions varied between 7.0 h and 30.8 h, in a Quad-Core processor with 4 GB of RAM.

The design of aviation networks under PSO is optimized in [6], minimizing both operational costs and total social costs. The objective of this article is to develop both demand predictive models and optimization models for flight scheduling and fleet planing, building on the model developed in [8] by adding the cost of connection time for passengers. The case study consists on two PSO networks: one based in Rhodes and another based on Thessaloniki.

3. Methodological Approach for Demand Prediction

Firstly, one needs to identify significant explanatory variables in order to define the demand prediction model. Then, it can be used to estimate the three different passenger demand scenarios to feed as input to the optimization model, as summarized in Figure 2.

3.1. Introduction to Gravity Models

Gravity models are suitable to address previously non-existent routes and deal with limited historical data. They can be formulated by

$$V_{ij} = k\frac{(A_iA_j)^\alpha}{d_{ij}^\gamma},\ i \neq j,\ d_{ij} = dji\ , \qquad (1)$$

where V_{ij} is the passenger volume between two cities i and j, A_i is the attraction factor for city i, d_{ij} is the distance between them, and k is a constant. Parameters α and γ control the influence of attraction factors and distance, respectively. Model (1) expresses non-directional passenger vol-

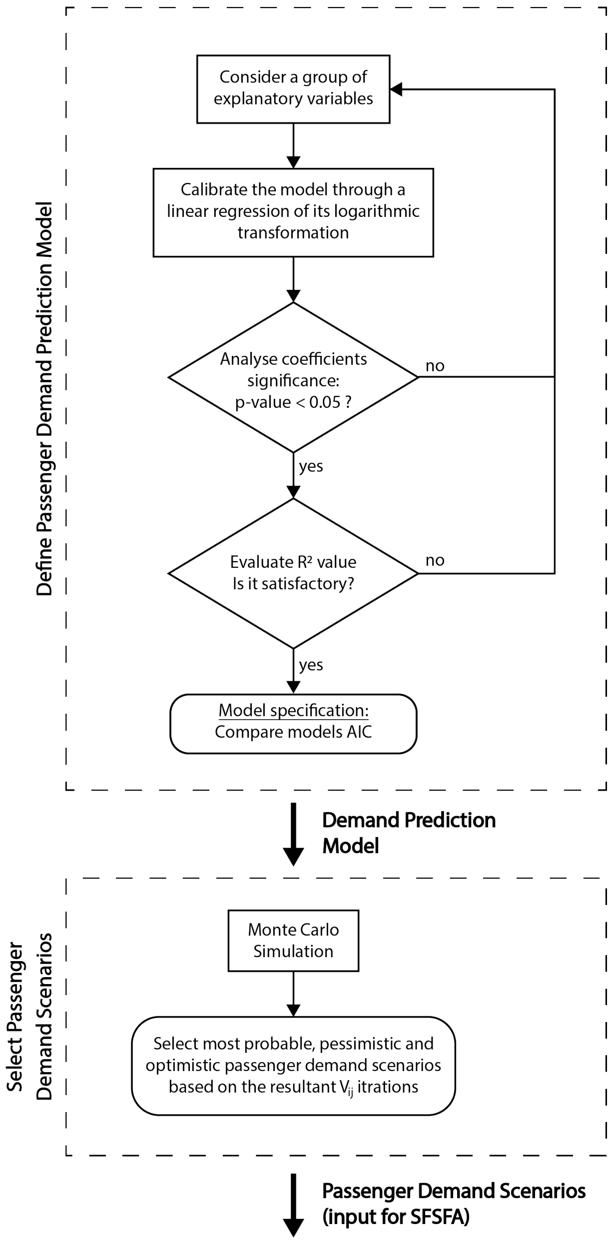


Figure 2: Methodological approach for passenger demand estimation diagram.

ume, but it can easily be adapted to a directional model by considering a deterrence factor B_i^{α} for the city of origin i and an attraction factor A_j^{β} for the city of destination j. Oftentimes, deterrence and attraction factors are the same, thus variables from (1) sustain, only allowing different parameters α and β for A_i and A_j, respectively [5].

Constant k and parameters α and γ can be estimated by linearizing (1) through a logarithmic transformation (base-10 logarithm is used),

$$\log(V_{ij}) = \log(k) + \alpha \log(A_i A_j) - \gamma \log(d_{ij}) \ , \quad (2)$$

and then performing a linear regression with data from already established routes using least-squares estimation. The attraction factor is usually expressed by a combination of explanatory variables.

3.2. Explanatory Variables

The assessment of the significance of each explanatory variable was based on the *p-value* of the null-hypothesis, which is that there is no relationship between the explanatory variable under discussion and the dependent variable. A 5% significance level was considered. The R-squared value was used to assess the absolute overall quality of the model. However, other concerns like the presence of multicollinearity and overfitting were considered. The Akaike Information Criteria (AIC) was calculated, as it is a comparative statistic measure that accounts for both their goodness of fit and their complexity, defined by

$$AIC = -2l + 2p, \quad (3)$$

where p is the number of estimated parameters (i.e. explanatory variables + intercept) and l stands for the maximum log-likelihood.

3.3. Design of Demand Scenarios

The model previously described is used to obtain three different passenger demand scenarios to use as input for the Route and Flight Schedule Optimization program. A Monte Carlo (MC) simulation, schematized in Figure 3, is performed to compute such estimations accounting for the covariances between the model's coefficients, and hence conferring robustness to the predictions.

For each iteration, the coefficients themselves are sampled assuming a multivariate normal distribution with the mean and covariance values that resulted from the regression of the logarithmic transformation presented in (2). The sampled coefficients are applied to the model to compute V_{ij} for a desired number of connections c, between airports i and j. Each V_{ij} observation is stored as a vector V of length c.

The criterion to stop iterating is based on the convergence of the trace of the sample covariance

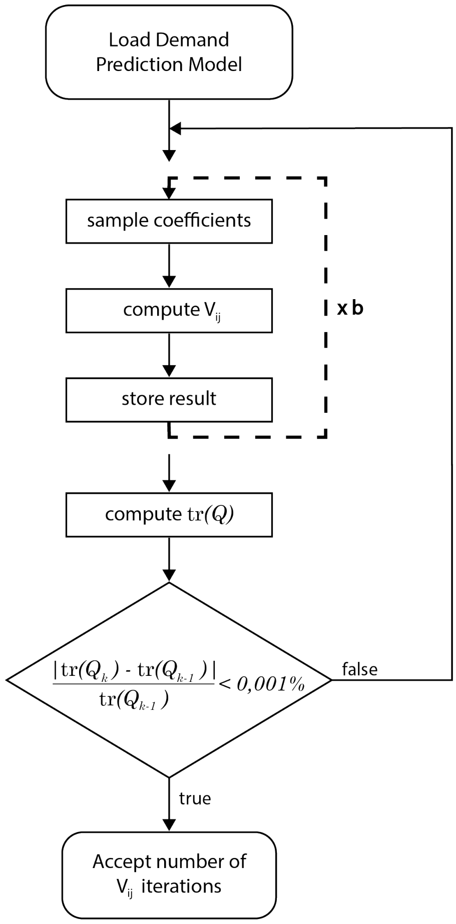


Figure 3: Monte Carlo simulation diagram.

matrix, Q. The three steps described above are repeated b times. At the end of each batch k of b iterations, $\mathrm{tr}(Q)$ is computed, with

$$Q = \frac{1}{N-1} \sum_{n=1}^{N} (V_n - \overline{V})(V_n - \overline{V})^T \quad , \quad (4)$$

where N stands for the number of observations, equal to $k \times b$. The sum of the c diagonal entries of Q_k, $\mathrm{tr}(Q_k)$, computed at the end of batch k, is compared with the result from the previous batch $k-1$. If the relative difference between $\mathrm{tr}(Q_k)$ and $\mathrm{tr}(Q_{k-1})$ is lower than 0.001%, then the number of iterations is considered sufficient. If not, another batch of iterations is performed.

The distribution of the resultant MC iterations, for the sum of passengers for every route, $\sum_{ij} V_{ij}$, is then used to define the most probable, a pessimistic, and an optimistic scenario, using the median and percentiles. $P_{V_{ij}}$ stands for a generic percentile from the distribution of the iterations for passengers for route ij, $P_{\sum}$ stands for a generic percentile from the distribution of the iterations for the sum of $\sum_{ij} V_{ij}$. There can be more than one iteration $m \in \{1, \ldots, M\}$ with the same total number of passengers $\sum_{ij} V_{ij_m}$, yet differently distributed by the routes (with V_{ij_m} standing for the number of passengers for route ij in iteration m). Therefore, the differences between the number of passengers estimated for route ij in iteration m and their individual percentiles, $|P_{V_{ij}} - V_{ij_m}|$, are taken into account. This way, the iteration m in a given percentile that presents the lowest sum of

$$\sum_{ij} (P_{V_{ij}} - V_{ij_m})^2 \tag{5}$$

is selected as the scenario referring to the said percentile.

4. Methodological Approach for Optimization

The contribution of the developed optimization model is the introduction of different scenarios of passenger demand in order to account for uncertainty. Therefore, the focus is not to enhance state-of-the-art models by considering detailed operation costs in order to get closer to a *real-world* application, aiming for a fully integrated model, but to explore robustness. With this in mind, a simple approach based on [6] was followed to be then further extended by assuming a robust optimization perspective and carefully designing different passenger demand scenarios to accomplish that. The only costs considered are: the costs for the airline related to the operation of flights and to parking aircraft for longer than 2 hours, and the social costs related to passengers time onboard and waiting on ground between flight legs.

The formulation of the mixed integer linear programming (MILP) problem includes the definition of the sets, parameters, decision variables, and constraints. For more details, see the full thesis.

To satisfy the demand for a link (combination of origin i and destination j), passengers are assigned to direct routes, or to itineraries with up to three stops, i.e. up to four flight legs. Operational time per day is segmented into smaller time periods, with t representing the initial instant of a time step or the period between time steps itself.

Decision Variables

$x(f, t, d, r)$: number of flights operating on route f, with aircraft of type r departing at time step t of day d;

$y(a, t, d, r)$: number of aircraft of type r on ground at airport a during time step t of day d;

$u_D(f, t, d, s)$: number of passengers placed on direct route f, departing at time t of day d, considering scenario s;

$u_1(g, t, d, w, s)$: number of passengers placed on the one stop itinerary g departing at time step t of day d, and waiting for w time periods in the stop airport, considering scenario s;

$u_2(h, t, d, w_1, w_2, s)$: number of passengers placed on the two stop itinerary h departing at time step t of day d, waiting for w_1 and w_2 time periods in the first and second stop airports, respectively, considering scenario s;

$u_3(i, t, d, w_1, w_2, w_3, s)$: number of passengers placed on the three stop itinerary i departing at time step t of day d, waiting for w_1, w_2, and w_3 time periods in the first, second, and third stop airports, respectively, considering scenario s;

$gc(a, t, d, r)$: binary decision variable that equals to 1 if aircraft of type r is on the ground at airport a for more than 2h as of time step t of day d.

Constraints

Fleet availability : limits the number of aircraft to the available fleet, by guarantying that the sum of aircraft of type r on the ground and flying corresponds to the available number of aircraft of that type r;

Time step continuity : ensures aircraft placement is coherent from one time step, $t-1$, to the next, t, i.e. for each aircraft type r, the number of aircraft on the ground during time period $t-1$ plus the aircraft arriving immediately before time step t has to be the same as the number of aircraft staying on the ground for time period t plus aircraft departing at instant t;

At hub maintenance : ensures that every aircraft begins and ends the operational day at the hub. Note that this constraint also guaranties continuity from one day to the next;

Seat capacity compliance : guaranties that, for every scenario, s, the number of passengers carried in each route f operated at time t of day d is lower or equal to the joint seat capacity of the aircraft flying that route. Note that one has to consider passengers carried in route f as a leg of their itineraries;

Pre-defined requirements compliance : ensures that the solution respects the minimum number of flights per week and minimum seats per week for route f, defined *a priori*;

Airport fee charging: sets binary decision variable gc to 1 whenever an aircraft stays on the ground for more than two hours (assuming each time period t is equivalent to half an hour).

Objective Function

The minimization problem follows

$$\text{minimize} \sum_{k=1}^{9} O_i \quad , \tag{6}$$

subject to the aforementioned constraints, where O_k, with $k = 1, ...9$ stand for:

Operating flights costs, O_1: sum of the cost of flying every route f with aircraft of type r flown at any time t during all operational days d;

Parking aircraft costs, O_2: sum of the cost of having aircraft of any type r parked for more than two hours at any time t during all operational days d;

Passengers onboard time costs: sum of passengers onboard time cost for direct routes f, O_3, and up to three stops itineraries g, h, and i, corresponding to O_4, O_5, and O_6, respectively;

Passengers waiting time costs: sum of passengers waiting on ground between legs time cost for up to three stops itineraries g, h, i, corresponding to O_7, O_8, and O_9, respectively.

Aiming to reduce the number of decision variables to be considered by the software, u_D, u_1, u_2, and u_3 are declared as dynamic arrays, so that each array cell is explicitly created only for possible itineraries. In this way, when running the loops performed for each constraint, only the existing entries of these arrays are considered, instead of enumerating all possible tuples of indices. Thus, this approach reduces computational effort by implicitly applying in every loop the "At hub maintenance" constraint, and keeping several stops itineraries comprised in a single operational day.

Problem Relaxation

This problem is too complex to be optimized within a reasonable time, since it is applied to an entire operational week with passenger demand on both ways for almost every combination between the 8 airports (56 links). Therefore, must be relaxed to ease tractability. A cluster strategy, where islands were separated into clusters and restrictions for inter-cluster traveling were defined, was pursued. This approach is detailed in Section 7.1, upon the application in the Greek islands case study in particular. This technique allows the number of possible itineraries to be significantly reduced, maintaining the problem fairly close to the original, as it is a reasonable approximation. The permitted direct routes and multiple stops itineraries are pre-computed and fed directly to the optimization software.

5. Rhodes based PSO Network Case Study

The region under study is the Dodecanese Archipelago (EL421), that comprises one big island, Rhodes, with 115,000 inhabitants (according to 2011 Census), surrounded by 7 smaller islands with airports (mean population of 10,236), and more than a hundred even smaller islands, only 18 of which are inhabited, that will not be accounted for. Figure 4 presents a map identifying all airports mentioned in the present work. For the sake of readability, every airport is referred to by its airport IATA code.

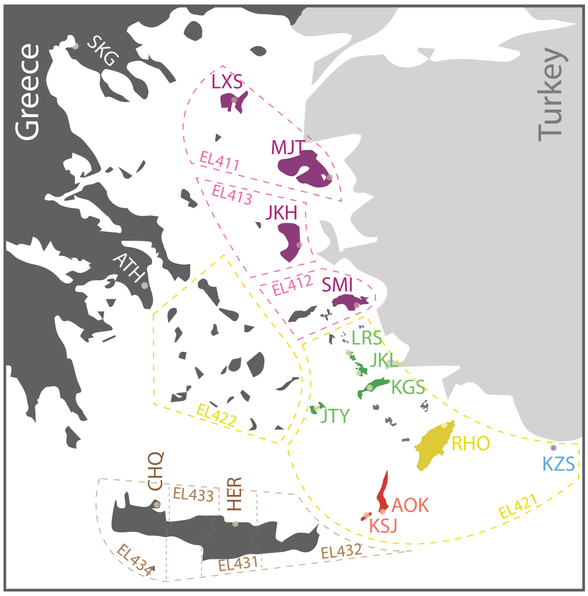


Figure 4: Case Study Greek Islands map.

The Greek islands to be analyzed within the present work case study are covered by PSO routes that link them with Rhodes Diagoras Airport (RHO). There are four "PSO groups", which are highlighted with color (this nomenclature will be used hereafter and RHO is considered a "PSO group" by itself, as it is the common airport for all four analyzed subsidized routes).

In the Nomenclature of Territorial Units for Statistics (NUTS) classification a higher level stands for smaller divisions within the previous level defined regions. The first, second, and third digits of the code stand for levels 1, 2, and 3, respectively. The network to be optimized by applying the method developed in the present work consists on the 8 islands within EL421 conferring an adequate complexity to the case study. As NUTS 3 correspond to the smaller divisions, the fact that many islands are located in a common NUTS 3 region can impair the usability of the available econometrics data.

All islands under discussion have, of course, one airport, and at least one port serving ferry boats, an important alternative travel mode. The sea transport between islands in Greece also relies on the governments' subsidies.

Tourism is an important contributor to the Greek economy. Eurostat points the number of nights spent at tourist accommodation establishments as a key indicator for analyzing the tourism sector, and the list of the EU regions with the highest numbers in 2019 is dominated by coastal regions around the Mediterranean Sea [4]. South Aegean islands

(EL421 + EL422) in particular have a great touristic presence: with a population of 340 870, only 3.17% of Greece's population, it registered 27 million nights spent at tourist accommodation establishments, according to Eurostat [1], about 25% of the overall total in Greece.

Still regarding tourism, seasonality is an essential factor. With 45.5% of total nights spent in tourist accommodation in 2016 occurring in only two summer months (July and August), making Greece the 2nd EU Member State with the highest seasonal deviation [3].

A crucial event to take under consideration in this case study is the Covid-19 pandemic, as it strongly affected not only passengers' mobility but the entire world economy, and, therefore, certainly tourism. In August 2020, Greece suffered a 67.68% decrease in nights spent at tourist accommodation establishments when compared to August 2019 [2].

6. Demand Prediction Model

6.1. Description of available data-set

The choice of the type of prediction model to develop was based on the available data. The lack of time-series data available hinders the application of Panel-Data Techniques or the introduction of any kind of dynamic component to the passenger demand model. The development of a gravity model was considered the most suitable approach, as literature suggests it is fit to predict demand on previously non-existing connections or when limited data is available.

When calibrating every model hereafter presented, data provided by airport infrastructure manager Fraport Greece was used for the dependent variable, passenger demand volume for connections between islands i and j, V_{ij}. Note that this data only included connections with i or j referring to RHO. There were 82 valid data points for V_{ij}, corresponding to the sum of passengers flying a link, throughout a month. Table 1 presents the main descriptive statistics of the sample. The table includes the explanatory variables used, with d for distance between airports (km), P for Population, G for Per Capita GDP, and T for number of nights spent by tourists in that island.

Table 1: Data-set descriptive statistics.

	V_{ij}	d_{ij}	P_i	P_j	T_i	T_j
Mean	463	206	107657	109332	230	235
Med.	123	154	115000	115000	296	296
SD	836	112	154206	153770	129	127
Min.	4	100	492	492	15	23
Max.	4090	465	623065	623065	562	562

6.2. Identification of Explanatory Variables

Several different combinations of explanatory factors were explored and two of them are presented hereafter. MATLAB was the software used to estimate the linear regression models throug a standard least-squares method.

Model 3

This model considers population of island i, P_i, a variable referring to the number of nights spent by tourists in island i in a year per capita, T_i, as a proxy for that island's touristic attractiveness, and a dummy variable, to distinguish high and low demand seasons, S, with $S = 10$ for August, and $S = 1$ for November:

$$V_{ij} = k\, d_{ij}^{\gamma}\, P_i^{\delta}\, P_j^{\epsilon}\, T_i^{\theta}\, T_j^{\iota}\, S^{\lambda} \,. \tag{7}$$

As the base-10 logarithm is used for the transformation, $S = 1$ (low season), is equivalent to not affecting the prediction, when $S = 10$ (high season), the prediction will be multiplied by the factor 10^{λ}.

The results of the linear regression of the logarithmic transformation of (7) are presented in Table 2. All added variables were shown to be significant. Coefficients for T_i and T_j, assume positive values, as one would expect. λ also takes a positive value, as August clearly presents higher passenger volumes V_{ij}. In this case study, the ferry boat becomes more preferable when islands are closer, making this alternative mode of transport an important competition for lower distances. This could explain why the distance coefficient γ, with a good level of significance, assumed a positive value.

With an R^2 of 0.47, this model is reasonably adapted to this case study but still suitable for other touristic regions with different characteristics. As shown is Table 3, when analyzing *Model 3*, the only correlation to be concerned about is the one between d_{ij} and T, but it can be explained by a mere coincidence of this particular case study, as the North Aegean islands are both less touristic and further away from Rhodes, consequently showing a negative correlation coefficient.

Model 7

The addition of a Covid-19 dummy variable resulted in a high p-value for the respective coefficient, which indicates that separate models would be more suitable to describe the years before and during the pandemic, as different factors may influence V_{ij} under such circumstances. However, this approach was not considered suitable because there is no sufficient data, only 55, and 27 data points for the periods previous and during the pandemic, respectively. Therefore, it was opted to consider the moderation of the explanatory variables according to the Covid-19 pandemic, resulting

$$V_{ij} = k\, d_{ij}^{C\gamma'}\, P_i^{\delta}\, P_j^{\epsilon}\, T_i^{(\theta+C\theta')}\, T_j^{(\iota+C\iota')}\, S^{\lambda}\, F_{ij}^{(\mu+C\mu')}\,, \quad (8)$$

where F_{ij} stands for a factor between ferry and air travel time, serving as a proxy for ferry competition, and $C = 0$ and $C = 1$ stand for the period before and during the pandemic, respectively. This variables were selected by performing a backward stepwise procedure.

The results of the linear regression of the logarithmic transformation of (8) are presented in Table 2. Tourism, T_i and T_j, continues to have a positive effect on passenger demand, both before the pandemic and during the pandemic, but less prominent during the pandemic period ($0 < \theta + \theta' < \theta$ and $0 < \iota + \iota' < \iota$). The same is verified for ferry competition F_{ij}, with coefficient $\mu + C\mu'$. Population, P_i and P_j, and seasonality, S, seem to influence similarly passenger demand, independently of the pandemic.

With $R^2 = 64\%$, *Model 7* is more tailored to this specific case study as they consider the ferry boat competition. When interpreting it, it is important to acknowledge the correlation between variables d_{ij} and F (presented in Table 3) when analyzing the effect of these variables on passenger demand V_{ij} within the pandemic context. The positive correlation between these two values was expected. However, *Model 7* variables are moderated with C, making the coefficients for the period before the pandemic ($C = 0$) equivalent to 0 for d_{ij} (γ disregarded) and 1.0 for F ($\mu = 1.0$), whereas for the period during the pandemic ($C = 1$) coefficients are equivalent to 0.76 for d_{ij} ($\gamma+\gamma' = 0+0.76$) and 0.28 for F ($\mu + \mu' = 1.0 - 0.72$), making this correlation irrelevant for estimations outside of the pandemic context, i.e. it is correct to interpret the coefficient $\mu = 1.0$ as the effect of ferry competitiveness on passenger demand when no pandemic influence is verified. Within the pandemic context, one cannot discriminate the effect of distance from the effect of ferry competitiveness on demand volume by analyzing the coefficients estimated for *Model 7*.

Table 2: *Models* coefficient estimates and p-values.

Coef.	Var.	*Model 3* (7) Est.	*Model 3* (7) p-value	*Model 7* (8) Est.	*Model 7* (8) p-value
$\log(k)$	-	-6.8	<0.01	-4.2	<0.01
γ	d_{ij}	1.5	<0.01	-	-
δ	P_i	0.17	0.03	0.20	<0.01
ϵ	P_j	0.19	0.01	0.22	<0.01
θ	T_i	0.91	<0.01	0.86	<0.01
ι	T_j	0.79	<0.01	0.75	<0.01
λ	S	0.38	<0.01	0.35	<0.01
μ	F_{ij}	-	-	1.0	<0.01
γ'	d_{ij}	-	-	0.76	0.010
θ'	T_i	-	-	-0.31	0.11
ι'	T_j	-	-	-0.27	0.17
μ'	F_{ij}	-	-	-0.72	0.02
R^2		0.47		0.64	
l		-50.63		-35.25	
p		7		11	
AIC		115.26		92.50	

Table 3: Correlation coefficients.

	$\log(d_{ij})$	$\log(P)$	$\log(T)$	$\log(S)$
$\log(d_{ij})$	-			
$\log(P)$	0.28**	-		
$\log(T)$	-0.42***	0.26**	-	
$\log(S)$	-0.12	0.04	0.07	-
$\log(F)$	0.55***	0.16	-0.2*	-0.05

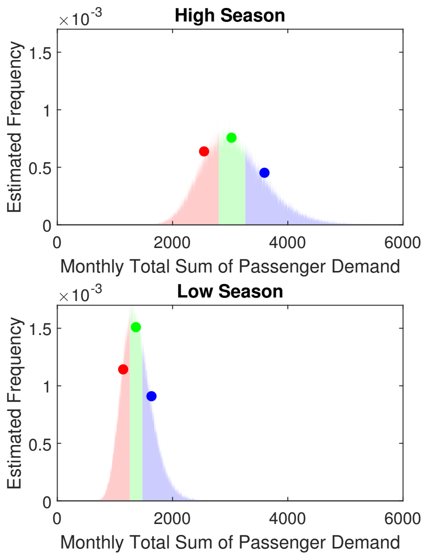
***p-value<0.01; **p-value<0.05; *p-value<0.10

6.3. Scenarios Estimation

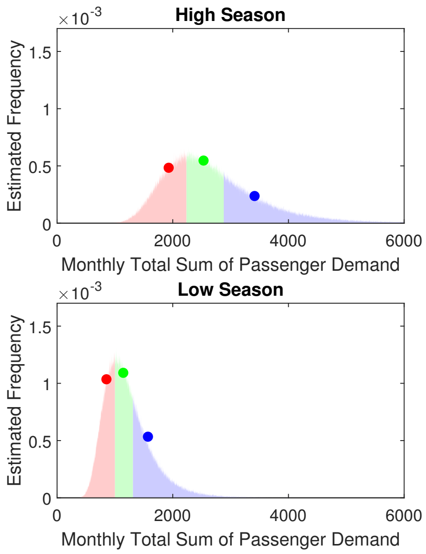
The data available to compute the models is inadequate for the application of the methods devised. In fact, given that only direct flights to and from RHO, which include passengers in multiple stops itineraries, are available, the predicted demand is grossly over-estimated. Therefore the results had to be normalized using an artificial constant. It was computed so that: i) the number of observed passengers (all traveling to and from RHO), and ii) the number of estimated passenger demand between airports from different PSO groups, are in the same order of magnitude. This rationale ensures that the number of passenger demand estimated for the whole network is not under-estimated either. If suitable data were, i.e. origin and final destination for each passenger, the devised methods could be used without further considerations.

When using *Model 7*, the moderation of the pandemic context effect in other explanatory variables was considered, therefore an additional step was added to the Monte Carlo simulation loop described in Figure 3, so that the probability of occurring pandemic context, p_C (probability of $C = 1$), was considered.

The most likely scenario assumes one of the iterations corresponding to the median of the total sum of passengers, highlighted in Figure 5 with green dots), selected according to the criterion described in Section 3.3. The percentiles used to obtain the pessimistic and optimistic passenger demand scenarios to feed as input to the optimization program, correspond to 1/6 and 5/6, and highlighted with red and blue dots, respectively. The rationale behind these percentiles is the split of the iterations in three equally likely ranges for the total sum of passengers (colored areas) and use their median to represent them (colored dots).



(a) outside pandemic context ($C = 0$)



(b) within pandemic context ($C = 1$)

Figure 5: Monthly total passenger demand estimated using *Model 7* iteration frequency, outside and within a pandemic context.

7. Implementation of the Optimization program

The optimization program will be applied considering time steps of half an hour, since air travel time between islands ranges between 23 and 61 minutes, corresponding to 1 to 3 time steps. This amount of time is reasonable because the smaller the time steps the higher the complexity of the optimization problem, which should be avoided, but a longer time step would impair the usability of the results, as most trips would be assumed to take the same travel time. The operational days are assumed to begin at 8 a.m. and end at 12 p.m., corresponding to 16 hours per day. It is beneficial to have the least number of operational days per week possible, as it allows a reduction of operational costs and also reduces the complexity of the optimization problem.

The higher the number of permitted waiting times are, the more inconvenient a several stop itinerary can be to passengers and the higher the complexity of the optimization problem is. However, if it is too low, then a great number of connection possibilities for non-direct itineraries is discarded, leading to an unfeasible problem.

7.1. Problem Relaxation

The clusters were defined accounting for the proximity between islands and the entry and exit airports were decided using an higher passenger volume criterion (Figure 6). Regarding the inter-cluster traveling restrictions, all clusters were allowed to connect directly with C_1, as RHO functions as a hub. Two more inter-cluster connections were allowed: i) between the two clusters with more islands, C_2 and C_4, in order to decrease the number of stops of non-direct itineraries, and ii) between C_3 and C_4 because of high passenger flow (Figure 7). Two additional restrictions were applied: i) multiple stop itineraries exiting and returning to the same cluster were excluded and ii) inter-cluster traveling within a multiple stop itinerary was limited to 3 clusters, i.e. only one intermediate cluster could be added besides the origin and destination cluster.

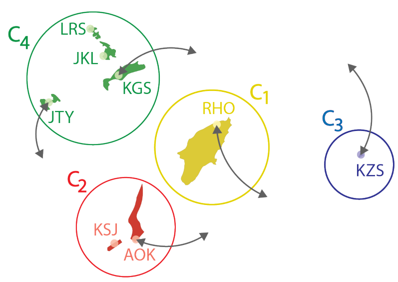


Figure 6: Entry and exit airports for each cluster.

8. Results

The results from the application of the Route and Flight Schedule Optimization program to the case study presented hereafter were obtained using the passenger demand scenarios predicted with

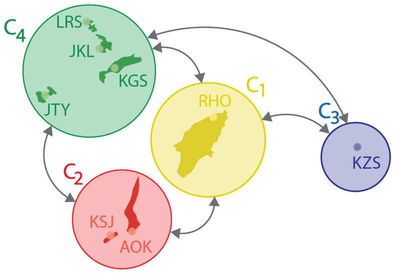


Figure 7: Inter-cluster traveling restrictions.

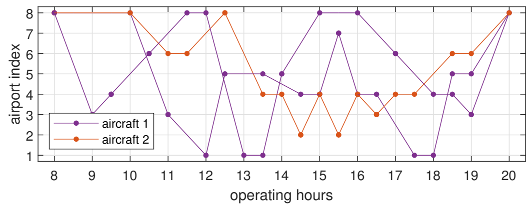


Figure 9: Solution for one day with 12 operational hours.

Model 3, on an Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz with 16 GB of RAM with the optimizer version 35.01.01 of the IBM's FICO Xpress software.

Figure 8 shows the evolution of the optimality gap considering two days with 12 operational hours each. Table 4 summarizes those results.

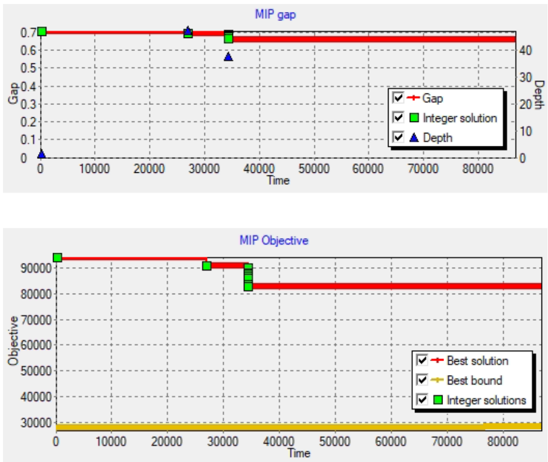


Figure 8: MIP search optimality gap, solutions depth, and objective evolution for two days with 12 operational hours each.

The allocation of the whole passenger volume to one single operational day, with the same 12h led to the solution presented in Figure 9. It assumed a 43.18% optimality gap, with 24 hours of computation time (decreasing insignificantly since the first hour and a half). On one hand, to fairly compare these networks, the cost of the three aircraft being parked on the hub airport RHO for one entire day must be added, resulting 45,021 €. On the other hand, significant savings due to closing the airport for the entire day would possibly overcome those extra expenses.

As expected, taking into consideration different passenger demand scenarios significantly increases the complexity of the problem. The deterministic approach that lead to the solution in Figure 10, reached a 32.50% optimality gap, lower than the 43.18% from the 3 scenario approach, in less computation time. The cost of this solution corresponded to 33,127 €, but this value should not be directly compared to the 3 scenario approach. The expected value must be calculated instead, as shown in Table 5. It was expected that the cost upon the verification of the most likely scenario would be lower for the deterministic approach, as it was optimizing the network precisely for that situation. However, this network would not be able to serve all passengers in case of the optimistic scenario, therefore, this network is not prepared for the uncertainty inherent to passenger demand.

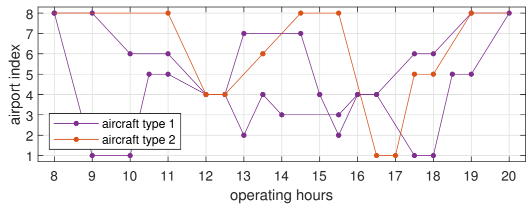


Figure 10: Solution for problem with passenger demand estimated with *Model 3*, cluster relaxation (b)(iii), no passenger partition, considering just the most likely scenario.

Table 4: Optimization results.

number of days	2	1	1
number of scen.	3	3	1
comp. time (s)	86400.0	99113.2	60412.3
opt. gap	65.90%	43.18%	32.50%
est. cost (€)	82663	42321	33127

The major conclusions one can draw from these results are that: i) decreasing the complexity of the problem (reducing the operational time and applying relaxation techniques) is essential to obtain better results and ii) although a deterministic approach reduces complexity it leads to solutions incapable to adapt to the uncertainty inherent to demand prediction. Therefore, the Route and Flight Schedule

Table 5: Calculation of estimated costs.

Costs		3 scen.	1 scen.
network operation	O_1	36799	27787
	O_2	300	450
passangers time	pessimistic	4130	3975
	m. likely	5055	4890
	optimistic	6480	unfeasible
est. total	$\sum_{k=1}^{9} O_k$	42321	-

Optimization program developed, considering several demand scenarios, was proven to be a more robust tool.

9. Conclusions

The major achievements of the present work consist on the methods developed to account for uncertainty, namely: i) the prediction of different passenger demand scenarios looking upon the covariances between the estimated prediction model's coefficients, ii) the development of the robust Route and Flight Schedule Optimization program, prepared to consider them simultaneously, and iii) the definition of relaxation techniques in order to handle the computational complexity of the problem.

The development of the Passenger Demand Prediction models for the Rhodes based PSO network case study, indicated the influence of the population of the islands, their touristic attractiveness, the seasonality, and the sea transport competition. It was also shown that the Covid-19 pandemic affected the impact of some of these explanatory variables. *Model 3* (that accounted for the effect of the population, the touristic attractiveness, and the impact of seasonality) is easily portable to other remote regions with similar characteristics to the explored case study. Examples of these characteristics are: i) the lack of surface transport alternatives (verified, for example, in islands), ii) routes with reduced attractiveness commercially-wise because of low passenger demand volumes, or iii) strong seasonal impact (e.g. due to tourism). The simulation methods applied to generate the pessimistic, most likely, and optimistic scenarios, taking the covariances between coefficients of the models into account, are also a novelty.

The consideration of several passenger demand scenarios when optimizing the network's routes and flight schedule is the major novelty presented in the present work. This approach allows the acknowledgment of the uncertainty not only pertaining the passenger demand prediction techniques, but also the partial arbitrariness of passenger behavior itself and exogenous factors, like the occurrence of unexpected events. The results have shown that, although considering only the most likely scenario may result in decreased costs when that scenario is verified, the solution from that deterministic approach would not be capable of serving a plausibly optimistic demand scenario. Therefore, the developed Route and Flight Schedule Optimization tool is proved to be robust and suitable to be integrated in a decision support framework regarding subsidized routes with low passenger demand.

The handling of the problem complexity through the definition of two different relaxation methods: cluster restrictions and passenger demand partition, is also an important contribution. It stood clear that the relaxation of the original problem was essential to obtain solutions in practical computational times. Albeit introducing a mild approximation, these relaxation approaches are shown to be vital to obtain better solutions in reasonable computational time than if no relaxations were employed.

References

[1] Nights spent at tourist accommodation establishments by nuts 2 regions. *Eurostat Data Browser*, accessed on 24.09.2021.

[2] Percentage change (compared to corresponding period of the previous year) in nights spent at tourist accommodation establishments by residents/non-residents - monthly data. *Eurostat Data Browser*, accessed on 24.09.2021.

[3] Research for TRAN Committee - Transport and Tourism in Greece, 16.4.2018. *European Parliament Website*, accessed on 24.09.2021.

[4] Tourism statistics at regional level. *Eurostat Statistics Explained*, accessed on 24.09.2021.

[5] T. Grosche, F. Rothlauf, and A. Heinzl. Gravity models for airline passenger volume estimation. *Journal of Air Transport Management*, 13(4):175–183, 2007.

[6] F. Leandro, A. R. Andrade, and S. Kalakou. Designing aviation networks under Public Service Obligations (PSO): A case study in Greece. *Journal of Air Transport Management*, 93:102042, 2021.

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

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