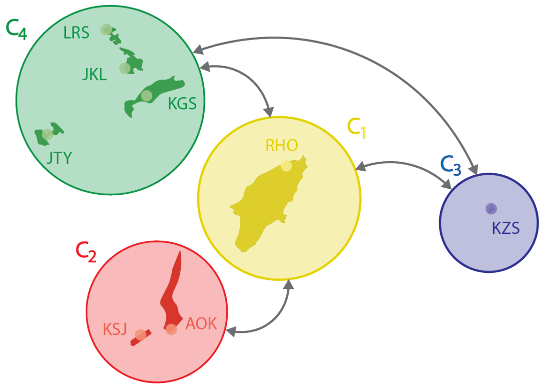


Robust Network Design for Air Transport under PSO

Alice da Silva Costa Lourenço

Thesis to obtain the Master of Science Degree in

Aerospace Engineering

Supervisors: Prof. António Ramos Andrade
Prof. Sophia Kalakou

Examination Committee

Chairperson: Prof. José Fernando Alves da Silva
Supervisor: Prof. António Ramos Andrade
Member of the Committee: Prof. Susana Margarida da Silva Vieira

June 2022

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

First, I would like to express my appreciation to my supervisors, Professor António Andrade and Professor Sophia Kalakou, for all the support and guidance through the development of this thesis. I sincerely thank you for being always ready to answer my questions, to discuss ideas, and to encourage me to go further.

I am deeply grateful that I have had the opportunity to study for the past five years and finally develop this thesis. I appreciate the Portuguese public education system and I am thankful for the financial support given by Câmara Municipal de Lagos. A special thanks to Associação Duarte Tarré for the financial support and, above all, for the companionship between fellow grantees and the advice and tireless support from the patrons.

Many thanks to Janeiro, Mariana, Gabi, and Maria, for every text, every trip, every smile, every bite, every laugh. Thank you Leo for your ideas, for your corrections, for all your support, and especially for your shoulder. A big thank you to my family, for the encouragement to grow in so many ways. For my mother, who thought me how to color in and outside the lines, and for my father, who created with me our own games to play.

Resumo

A acessibilidade é essencial para o desenvolvimento social e económico e para a coesão entre regiões. Neste sentido, o transporte aéreo é indispensável, especialmente em zonas geograficamente remotas, como ilhas, mesmo quando as rotas são pouco rentáveis devido a baixa procura por parte de passageiros. As Obrigações de Serviço Público (OSP) visam preservar precisamente essas rotas, garantindo serviços de transporte mínimos às populações que servem.

O objetivo deste trabalho é desenvolver ferramentas robustas que possam apoiar as autoridades responsáveis a tomar decisões sobre estas redes subsidiadas. Neste sentido, um modelo de previsão de procura e um programa de otimização robusta de rotas e horários de voo foram desenvolvidos. Adicionalmente, dada a complexidade deste problema de otimização, foram apresentadas formas de relaxar o problema.

Estes métodos foram aplicados à rede de OSP centrada em Rodos (Grécia). Os modelos de previsão indicaram que a procura por parte de passageiros nesta rede é influenciada pela população de cada ilha, a atratividade turística, e a competição do transporte marítimo. Foi também verificado que a pandemia Covid-19 alterou o impacto de algumas destas variáveis explicativas. Foi ainda demonstrado que otimizar a rede tendo apenas em consideração o cenário mais provável resulta numa solução menos robusta, incapaz de servir os passageiros num cenário otimista.

Este trabalho contribui para o estado-da-arte ao considerar diferentes cenários de procura simultaneamente para a otimização. O que se mostra relevante, pois considerar apenas o cenário esperado, como proposto na literatura, não se adapta à incerteza inerente à previsão da procura.

Palavras-chave: Obrigações de Serviço Público; Ilhas Gregas; Otimização Robusta; Desenho de redes de transporte aéreo; Previsão de Procura de Passageiros

Abstract

Accessibility is essential for social and economic development and cohesion of regions. Thus, air transport is vital, 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 unprofitable yet vital routes, assuring that minimum transportation services are granted to those remote 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. Furthermore, given the computational complexity of this optimization problem, relaxation approaches were introduced.

To validate the proposed approach, it 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. This consideration is shown to be vital, since the optimization taking the expected demand solely into consideration, which is proposed in the literature, 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

Contents

List of Tables

List of Figures

List of Acronyms

AIC Akaike Information Criterion

EU European Union

FSFA Flight Scheduling and Fleet Assignment

GDP Gross Domestic Product

GVA Gross Value Added

IFSFA Integrated FSFA

MC Monte Carlo

MILP Mixed Integer Linear Programming

NUTS Nomenclature of Territorial Units for Statistics

PSO Public Service Obligation

SFSFA Socially-oriented FSFA

TRAN Committee on Transport and Tourism

Chapter 1

Introduction

Accessibility is a main driver of economic development (Smyth et al. [1]). Therefore, air transport is an industry of strategic relevance both at international and regional levels. On the one hand, it takes an important role on the governments' international relations policy. On the other hand, air transport comes as absolutely essential for social and economic development and cohesion of regions, especially in geographically remote areas, like islands, where there are not many surface transport alternatives [1].

In Europe, the process of air transport liberalization was part of the Single European Act of 1986, aiming to form a single internal market across a whole range of economic activities. This was a long process, as several sets of measures were gradually implemented, up to the 'third package', in 1992. All decisions on the 'European Single Aviation Market' had and have to be thoroughly studied, given the significance of their consequences. One of the main concerns 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: according to Zhang and Graham [2], the relationship between air transport demand and the economy is bi-directional, being prevalent in less developed economies, which consequently rises a vicious cycle that amplifies disparities by not contributing to the financial stability of remote areas.

It was then deemed necessary to create means of preserving vital yet unprofitable routes, assuring that minimum transportation services were granted to remote communities. It is essential to ensure that the residents have access to a transport system that confers connectivity to main urban areas, preventing an inhibition of movement that would keep families apart, and suppress business and tourism growth [1]. The solution found was to create a subsidy scheme that gives those smaller economies a chance to grow, while assuring that a fair competition basis is kept in the market and that the financial support given does not turn out as a burden for taxpayers.

The European Union (EU) defined Public Service Obligations (PSOs), that 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 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.

1.1 Motivation

The important effect of the application of PSOs in routes with low demand is unquestionable, as it affects not only the daily lives of the people from the region they serve and their opportunities, but often the whole country's economy, for instance the tourism sector. Naturally, every decision on this topic is embedded in a broader strategic plan, which makes the problem much more complex, as there are different parties involved, with different objectives to be taken into account. It should be also noted that transparency is key, since handling public funds always raises debates about efficiency in public policy and real contribution to society, thus it is important to follow clear and pragmatic criteria. The objective of this work is not to deeply study the decision process as a whole, but to focus on some of the necessary tools to improve that process, as market imperfections can impede success in securing adequate air transport to facilitate economic development [2].

It is imperative that the responsible authorities use every tool at their power to support decisions on what routes to subsidize and what standard level of service to set in the PSO public tender. These decisions should be based on robust assumptions, that take the inevitable uncertainty associated with passenger demand prediction into account.

1.2 Topic Overview

To better understand the topic, one has to be aware of PSO legal framework and general principles. Moreover, one has to understand who are the stakeholders involved and what are their objectives, enabling to conceive a method to support public authorities to formulate and decide on the standards required in a public tender in order to deliver PSO routes that serve the needs of both citizens and airlines. This is followed by a literature review on the state-of-the-art existing tools needed to carry out that method, as those are the focus of this work.

1.2.1 PSO Framework

Instead of enumerating exhaustively all rules and guidelines defined by the EU, only what is relevant to support the developed work is presented hereafter. According to Regulation (EC) No 1008/2008 of the European Parliament and of the council [3], *"The necessity and the adequacy of an envisaged public service obligation shall be assessed by the Member State(s) having regard to:*

- *the proportionality between the envisaged obligation and the economic development needs of the region concerned;*

- *the possibility of having recourse to other modes of transport and the ability of such modes to meet the transport needs under consideration (...);*
- *the air fares and conditions which can be quoted to users;*
- *the combined effect of all air carriers operating or intending to operate on the route."*

There are two types of PSO: Open and Restricted. The intention is to ensure that competition in the context of a PSO takes place to the widest possible extent, therefore only if there are no airlines interested in operating the envisaged PSO route as an Open PSO, the Member State can consider a Restricted PSO. In that case, the access to the scheduled air services on that route is limited to only one air carrier for a period of up to four years (or five, in outermost regions), after which the situation shall be reviewed. If exclusivity is insufficient to ensure the viability of service, then compensation can be awarded. The parameters to calculate such compensation must be clearly set in the published invitation to tender and subsequent contract and it may not exceed the net costs incurred in discharging each PSO.

The regulation allows to issue a public tender for a group of PSO routes, but only when justified for reasons of operational efficiency. There is also the option to define routes with one or more stopovers. Even so, *"the assessment of the adequacy of the PSO needs to be made for each flight segment individually"* - Commission notice on Interpretative guidelines on Regulation (EC) No 1008/2008 [4].

Each state member can define standards of continuity, regularity, pricing or minimum capacity as they see fit. Continuity obligations are particularly useful when a strong seasonal pattern is verified, so that minimum service is ensured on low demand periods. Regularity and capacity obligations mainly include the fixing of minimum capacities in terms of seats offered or of minimum frequencies. Maximum prices or tariff grids (to account for categories of passengers, e.g. residents and students) may be defined as well. It is also possible to define aircraft requirements, but these have to be carefully justified, as it is preferable to avoid restrictions that could lead to arbitrarily excluding specific air carriers. All requirements must be non-discriminatory and proportional to the economic and social development needs of the region concerned.

1.2.2 Decision Support Framework

The objective of creating PSOs, in the first place, is to satisfy passengers' needs, but it is equally important to analyze the objectives of all the stakeholders involved. Passengers seek affordability (lower airfares) and convenience (higher frequency and more direct routes). Producers - airports and airlines - aim to minimize operational costs. Air Transport Public Authorities aim to ensure the viability of the network and the continuity of the service.

Even though the European Union PSO guidelines encourage to set less constrained requirements, allowing airlines to define their own flight schedules and fleet assignment could lead to the minimization of the cost incurred by the airlines but not necessarily the total cost, as it can potentially lead to passenger welfare worsening (referred to as social costs). The implementation of a robust Flight Scheduling and

Fleet Assignment (FSFA) solution for a whole region that maximizes passenger welfare and minimizes operational costs would justify tendering a group of PSO routes and narrowing the PSO requirements as it would be beneficial to all stakeholders. Pita et al. [5] and Leandro et al. [6] followed this approach by designing a Socially-oriented FSFA (SFSFA) optimization model, that will be discussed in Section 3.2.

The SFSFA tool would receive the passenger demand prediction and some other constants (needed to compute airline and social costs) as input and retrieve the optimal route and flight schedule for the network, given that data. It could be incorporated into a decision-making mechanism like the one schematized in Figure 1.1.

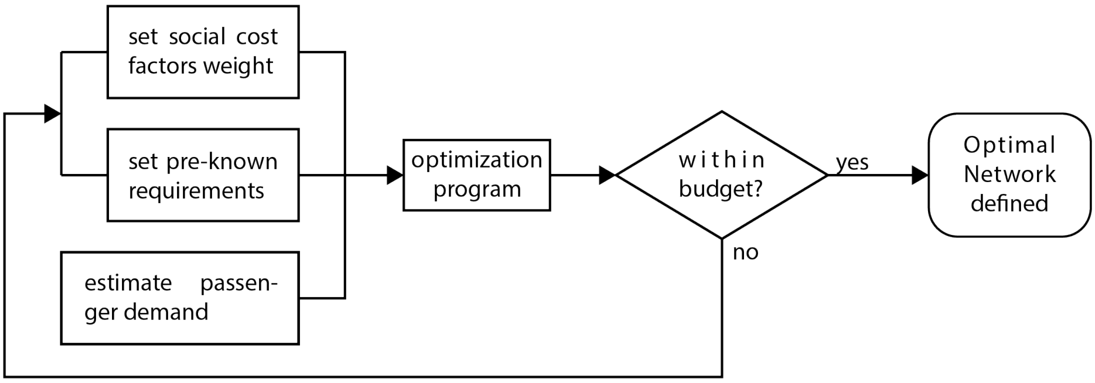


Figure 1.1: Example for a decision support framework.

Firstly, the weight to attribute to the different passengers' preferences would have to be determined, for example the travel time cost for passengers. If any pre-known set of requirements was to be defined, for example the need to connect a remote island with the closer health facilities, it should also be fed to the SFSFA as an input. Additionally, estimates of passenger demand are an essential input to SFSFA. Once the optimal solution is computed, taking every stakeholder's cost into account, the authorities have to verify if the necessary compensation to implement complies with the budgetary target. If it does, then the standards to require on tender should be determined so that the optimal network is actually implemented. If not, the process would have to be reiterated, adjusting the weights attributed to social cost factors and revising the pre-defined requirements.

1.2.3 Subjects of interest

To carry out the iterative mechanism described above, one has to apply methods of passenger demand prediction, as those estimations are needed as an input to the optimization problem. A robust network optimization method, adapted to the specific case study, has to be created as well.

Passenger demand estimation

To begin with, it is important to clarify that the objective of this step is to estimate the attractiveness of links between airports serving the region under study and not to predict the future passenger flows on currently available routes. This estimation aims to represent latent demand, which expresses the passenger's desire to travel a route regardless of the level of service (thus including non-existent routes), also called unconstrained demand. Various methods to estimate air travel passenger demand were studied. First, Panel-data Techniques were considered, as they allow for a comprehensive analysis of data on both time-series and cross-sectional aspects. Second, Gravity Models were explored as they were often suggested in the literature as a suitable solution to handle previously non-existent routes and lack of historical information.

Flight Scheduling and Fleet Assignment

Air transport operations have particularly low profit margins, therefore operational efficiency comes as absolutely essential. For this reason, there is extensive literature on FSFA from the airlines' point of view, i.e. aiming to maximize profits. Several different approaches can be taken, either trying to approximate the models to the real problem, adding complexity to the model, or relaxing it in order to simplify it for the sake of tractability. The main references for this dissertation are Leandro et al. [6], which in turn refers to Pita et al. [7] and [5], as they focus precisely on PSO standard definition, taking social costs into account in their optimization models. The multi-objective optimization is approached with the weighted sum method.

1.3 Objectives

The main goal of this work is to show the importance of considering passenger demand uncertainty when applying the demand prediction and SFSFA tools within a decision support method regarding PSO route and flight schedule definition. A common type of remote regions in need for subsidized air transport to ensure connectivity is islands, as there are not many surface transport alternatives and the routes may not be attractive commercially-wise because of low passenger demand volumes. The case study was chosen with these characteristics in mind. The Greek islands comply with this description and are also characterized by an important seasonal factor because of tourism, which can lead to further studies about the necessity of a different route and flight schedule network for high and low seasons, consisting on a relevant case study.

The first objective is to define a suitable demand prediction model for the PSO network in Greece that connects several islands with Rhodes as the hub. This is a challenging task because of the particular characteristics of this case study. First of all, there is no available data on demand for non-direct itineraries, but only for passenger flows between airports. There is, obviously, no passenger demand data for previously non-existing routes either. Secondly, as many islands belong to the same Nomenclature of Territorial Units for Statistics (NUTS) 3 region, the lack of econometrics data for each island comes as an additional constraint. Other singular aspects to take into account in this case study are the important seasonal interference, the sea ferry competition, and the Covid-19 pandemic context impact. Although passenger demand prediction is a widely discussed subject, a model handling such thin demand routes with the set of characteristics enumerated above comes as an interesting novelty, adding to the state-of-the-art existing literature.

The second part of this dissertation focuses on creating a robust optimization tool, based on SFSFA methods, which accounts for the uncertainty intrinsic to the demand forecast. As many authors developed the idea of what costs should be included in the SFSFA, in order to achieve an integrated solution, to the best of our knowledge, passenger demand uncertainty is not taken into account.

Different demand scenarios will be taken from the prediction model previously obtained by considering the coefficients' covariance. The optimization program will be prepared to consider those scenarios, fulfilling the constraints for all of them and optimizing costs by taking their probabilities into account. As the cost of the optimal solution is expected to increase for the most likely demand scenario, it is expected to better serve scenarios that plausibly differ from the most likely one. Results will be compared in order to show that this holds.

This work aims to contribute to the knowledge about what factors influence air transport passenger volume in this case study in particular and how to apply them to a suitable prediction model and to create a robust Route and Flight Schedule Optimization program applicable to different subsidized air transport networks. As a result, the contribution of this dissertation consists in the acknowledgment that uncertainty is highly relevant, and therefore it is essential to include it in the decision process.

1.4 Thesis Outline

In Chapter 1 a brief introduction to the context of the present work is presented, expressing the motivation to develop it, providing a topic overview, and finally specifying the objectives of this research, and briefly presenting the case study to which the developed method will be applied.

Chapter 2 explores some of the literature reviewed in order to acknowledge the state-of-the-art of the topics necessary to develop the present work, namely passenger demand prediction (Section 2.1) and operational optimization (Section 2.2).

In Chapter 3 the methodological approach to develop the present work as a whole is thoroughly explained. It starts with the estimation of passenger demand in Section 3.1, including the development of the prediction model and the definition of different scenarios. Then, in Section 3.2, the optimization problem is defined and the relaxation of the problem, in order to keep it tractable upon its application to the Greek Islands case study, is briefly explained. The particular characteristics of this case study are explained in Chapter 4.

In Chapters 5 and 6 the application of the previously explained methods to the case study is described. Chapter 5 begins by describing the available data-set, proceeds with the development of the demand prediction models, and finally presents the resultant passenger demand scenarios estimated. Chapter 6 explains the application of the optimization program to the case study, considering the passenger demand scenarios previously estimated, describing the challenges related to the tractability of the problem that emerged throughout it.

In Chapter 7 some of the previously presented intermediary results are discussed, namely the results from applying the prediction demand models. Some different combinations of relaxation methods and operational time periods to consider in the optimization problem are tested and thoroughly discussed by comparing their results. The uncertainty approach explored in the present work will also be compared with a deterministic approach, in terms of tractability and final cost results.

Finally, Chapter 8 summarizes the achievements and limitations of the present work and leaves some suggestions for future work on this topic.

Chapter 2

State-of-the-art

In this chapter, the state-of-the-art on air passenger demand prediction and on network design optimization (mentioned in Section 1.2.3) is explored. First, Section 2.1 focuses on the passenger demand and reviews research works on the application of different methods to estimate it. Then, Section 2.2 explores the previous research on air network design and operational optimization.

2.1 Passenger Demand Prediction

This subject is key to the air transport industry, as it is necessary to correctly implement FSFA methods. Different methods can be used to obtain passenger demand prediction models depending on the data available and the objective of developing those models. The different passenger demand prediction models covered are grouped according to their methods: i) multivariable linear regression; ii) panel-data techniques; and iii) gravity models. An additional section covers other interesting literature relevant to this case study.

2.1.1 Multivariable Linear Regression

Multivariable linear regression models are widely used. This method consists of estimating a linear relationship between the dependent variable (passenger demand) and each independent variable (two or more). Devoto et al. [8] examine the potential of econometrics for air transport demand forecasting. The proposed model uses both historical traffic data and socio-economic variables chosen using Student's t-test to assess statistical significance and examining the correlation matrix to analyze multicollinearity. As the case study was the region of Sardinia, a tourist destination, the socio-economic variables include the number of tourist beds and tourist arrivals. Note the similarities with the present work case study, as both refer to touristic islands in the Mediterranean Sea.

2.1.2 Panel-Data Techniques

Panel-Data Techniques allow for a comprehensive analysis of data on both time-series and cross-sectional aspects. Chevallier et al. [9] estimate the worldwide air traffic using econometric dynamic panel-data modeling. Their objective was to forecast passenger demand in the mid-term (over 15 years). Gravity models are pointed as the most intuitive approach, yet unsuitable for air traffic modeling at the worldwide level. The research points out three types of potential air traffic determinants: i) Gross Domestic Product (GDP) growth rates; ii) ticket price, utilizing jet-fuel prices as one of its main components; and iii) exogenous shocks. It additionally suggests that the weight of each of these determinants depends on the market maturity. The panel-data sample used is closer to time-series data (yearly data from 1980-2007) than cross-sectional data (8 geographical regions), thus a lagged dependent variable among the regressors was considered appropriate, conferring a dynamic structure to the model.

Tsekeris [10] used data from 1968 to 2000 to design a dynamic demand model for 18 flight connections from Athens airport (Attica) to 7 Aegean islands prefectures in Greece. The proposed dynamic panel-data model, by exploring both cross-sectional and time-series components, enables the relaxation of limiting assumptions, like the endogeneity of explanatory variables and the heterogeneity of travel behavior across destinations, commonly present in air travel demand prediction models. Another interesting aspect to note is that competition is also taken into account in this paper through the inclusion of air transport *versus* sea transport relative travel cost. This last aspect is interesting for the case study of the present work, as the ferry boats are an important alternative travel mode, as mentioned in Section 4.

2.1.3 Gravity Models

Gravity models consist on multiplying the origin and destination attractiveness factors, moderated by raising them to the estimated coefficient power, in order to estimate the demand between those airports/cities. Nõmmik and Kukemelk [11] present a gravity model for regional route planning and suggest methods of calibration. The model is applied to a set of 11 European Airports under some specific conditions: regularity of service, international routes under the Great circle distance of 1852 km) and not being exclusively served by low cost carriers. This paper emphasizes the convenience of gravity models, specially as an alternative in case of a lack of data on historical booking figures. It also states that gravity models are suitable to address regional and peripheral airports, as they are not distorted by multi-stop journey numbers and by travel with a combination of other modes of transport. Additionally, some interesting factors to describe a region's attractiveness are mentioned: the number of UNESCO World Heritage sites and the number of tourist nights spent per fixed period.

Wadud [12] studies passenger demand modeling and forecasting in the context of limited data. A gravity-type demand model for Bangladesh is applied to a new airport in a divisional capital as the case study. Panel-data, with time-series of air passenger demand and its explanatory factors, is pointed as the ideal data for modeling air transport demand. Since those were not available, national GDP and population are used as proxies at the cost of introducing some bias.

Grosche et al. [13] presents two gravity models, one limited to city-pairs with airports not subject to competition from airports in the vicinity, and other including all city-pairs, by using booking data of flights between Germany and 28 European countries to calibrate them. In this article, explanatory variables are categorized as geo-economic or service-related. Service-related factors include quality (for example, travel time, frequency of flights, airline's market presence) and price. Geo-economic factors account for economic activity (income, population served by the airport, historical passenger volumes at each airport, among others) and geographical characteristics (distance between origin and destination). The proposed models include only geo-economic factors, stating that they are suitable for links where currently no air service is established, historical data is not available, or for which factors describing the current service level of air transportation are not accessible or accurately predictable. This article also presents a summary list of previous studies identifying the different factors considered in those estimated gravity models, including preliminary studies, e.g. Doganis [14], that only uses distance and an aggregate activity-related measure, the historical passenger volumes at each airport, as explanatory variables.

Kopsch [15] analyses the demand for domestic air travel in Sweden. An interesting addition to the usual explanatory variables used for air travel demand prediction is the alternative modes of transport, in that case, rail and road. Another particularity of this research is the division between business and leisure travelers, that confers robustness to the designed model. Time series analysis is used and logarithmic transformations allow the interpretation of coefficients as elasticity.

2.1.4 Other Interesting subjects

Sambracos and Rigas [16] study the passenger reactions to the Greek islands' market deregulation. This article comprises analyses of both air and sea transport and the relation between them. Three surveys were performed in the years 1996 (before deregulation), 2000 and 2005, showing that air and sea modes are preferred by different market segments. Air traveling offers a better level of service but higher fares. In addition to being cheaper, boat traveling can be considered a vacation experience and allows passengers to transport their private vehicle. Accordingly, it was concluded that ferries attract more cost-conscious passengers (leisure passengers, aged 18-30 years old, traveling in groups) whereas air transport attracts more time-sensitive and higher-income groups (business travelers, traveling alone, aged over 50 years old). It showed that seasonality affects passengers' profile, as sea travel is more attractive in the summer because of its touristic activity nature and less attractive in the winter because of bad weather making sea traveling uncomfortable or even unsafe.

In an attempt to address the Covid-19 pandemic, as an unexpected event that undeniably affected air transport demand, the Ito and Lee [17] paper on the impact of the 2001 September 11 terrorist attacks on U.S. airline demand was analyzed. The paper aims to disentangle macroeconomic effects from the direct effects of the event. In summary, two components of the effect of this tragic event were separated: i) a transitory temporary shock of more than 30% and ii) an ongoing negative demand shift of approximately 7.4% yet to dissipate two years later. One could draw a parallel with the 2020 pandemic, as it caused strong socio-economic changes that affect transport demand in a broader sense and also led to the implementation of lockdown measures that directly restricted passenger mobility.

A summary of the most relevant references of the present work is presented in Table 2.1.

Table 2.1: Reference summary concerning Passenger Demand Prediction.

Study	Method	Explanatory variables
Devoto et al. [8]	multivariate linear regression	Historical traffic data, socio-economic variables (including number of tourist beds and tourist arrivals), and distance.
Chevallier et al. [9]	panel-data	GDP growth-rates, ticket price, and exogenous shocks.
Tsekeris [10]	panel-data	I.a., ratio of air to sea transport fare levels, ratio of the total nº of flights to ferry itineraries, ratio of aircraft to vessel seats, ratio of air to sea ferry travel times, and nº of beds for visitor accommodation.
Nõmmik and Kukemelk [11]	gravity model	I.a., nº of UNESCO World Heritage sites and number of tourist nights spent per fixed period.
Grosche et al. [13]	gravity model	Population, catchment, buying power index, GDP, distance, and average travel time
Doganis [14]	gravity model	Historical passenger volumes at each airport.
Sambracos and Rigas [16]	—	Tourism, seasonality, and ferry boat competition for this specific case study

2.2 Operational Optimization

As stated previously, there is extensive literature on operations planning, which can cover flight scheduling, fleet assignment, route design, and/or crew scheduling, from airlines' point of view. Although the majority of that work focuses only on maximizing airlines' profits which is not aligned with the objectives of a low demand subsidized network case study, it is still worth analyzing, as many interesting ideas and methods can be applied. Literature specific to subsidized schemes was also explored. Note that many of the references presented hereafter include passenger demand prediction models as well, as they are oftentimes needed to implement the optimization model under study.

2.2.1 Flight Scheduling and Fleet Assignment

Dožić and Kalić [18] develop a three-stage airline fleet planning model: i) fleet mix approximation, ii) fleet size determination, and iii) aircraft type selection. The first stage is based on a fuzzy logic system that takes passenger demand and distance between airports as inputs to roughly estimate the aircraft size needed to serve a given route network. Secondly, an heuristic algorithm is developed in order to determine the minimal number of aircraft needed, taking the results from the previous step into account (approaches both for short/mid-term planning, when flight schedule is known, and long-term planning are presented). The third stage consists of a multi-criteria decision process.

Iliopoulou et al. [19] investigate the design of a seaplane network connecting the Greek islands with the mainland as an alternative to the currently offered low quality ferry service and costly air transport, that can only serve islands with airports. This network's main objective is to improve connectivity with the mainland, but also considers transportation between neighboring islands. It proposes a multi-objective route planning model, solved with a meta-heuristic (genetic) algorithm, widely used in combinatorial optimization problems.

2.2.2 Subsidized Networks Optimization

Kinene [20] analyzed decision models for the procurement of subsidized air services, with application to the Sweden PSO network, where the PSO context and tendering process are thoroughly explained. Similar to the present work, the aim of the thesis is to design decision support models for transportation authorities to select the routes to subsidize and to define the appropriate requirements to set on the tender. However, the approach to this problem is perceptibly different. Firstly, the identification of regions to benefit from the subsidy scheme, through accessibility criteria, is addressed originating a first model. Then, the behavior of airlines when preparing bids is replicated with an optimization model that accounts for the tender requirements and minimizes subsidies while maintaining a reasonable profit threshold. Airlines are assumed to minimize subsidies as the authorities will select winners based only on price. Then, a third model to support the determination of winners is developed, considering the possibility of bids on route bundles. Two papers further describing this work were published, the first addressing the choice of subsidized routes [21]. This paper focuses on the formulation of the first model, which includes a passenger estimation model, and deeply analyzes the improved accessibility of the proposed network when compared to the current one.

Pita et al. [7] proposes an operational decision approach to support the definition of PSO requirements consistent with government budgetary target and applies it to the Azores PSO network. Based on traditional FSFA, an Integrated FSFA (IFSFA) is developed, which, instead of maximizing profit, minimizes total costs accounting for the interests of passengers (social costs), airlines, and government. Note that multivariable regression analysis was used for passenger demand estimation. Pita et al. [5] builds up on the previous paper by including airport costs and applying it to a PSO network in Norway. The designated 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 for the different scenarios considered varied between 7.0 h and 30.8 h, in a Quad-Core processor with 4 GB of RAM. Relying on this modeling work, Antunes et al. [22] propose significant adaptations in order to tackle the Azores PSO network case study with an approach closer to a real-world application. Working closely with SATA, the airline operating the subsidized network, a whole year of operations, distinguishing the days of the week and three seasons (peak, median, and low) is considered, also taking airport infrastructure into account.

Leandro et al. [6] focuses on optimizing the design of aviation networks under PSO, 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 planning, building on the model developed in Pita et al. [7] 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. This work proves itself useful, as the application of the optimization model led to reduced total network costs, regardless of the demand prediction model applied.

The most relevant references for the second component of the present work are Leandro et al. [6] and Pita et al. [5], as it builds on the optimization model presented by Leandro et al. [6] by accounting for uncertainty when considering several different demand scenarios simultaneously.

Chapter 3

Methodological Approach

This chapter details the methodological approach and methods to obtain both the passenger demand prediction model and the Route and Flight Schedule Optimization program, in Sections 3.1 and 3.2, respectively. Both sections specify the predictive and optimization models from Chapter 2 that most directly led to the models to be applied to the case study in Chapters 5 and 6.

3.1 Passenger Demand Estimation

From the different types of demand prediction models covered in the previous chapter, gravity models were the chosen approach to pursue, as they are the most suitable to address previously non-existent routes and deal with limited historical data. This choice will be more deeply discussed in Section 5.1.

Firstly, one needs to identify significant explanatory variables in order to define the model. Then, it can be used to estimate the three different passenger demand scenarios to feed as input to the optimization model. Figure 3.1 summarizes this section, from the passenger demand model definition to the different scenarios estimation.

3.1.1 Introduction to Gravity Models

Gravity models are inspired by the idea of gravity between two objects, which is directly proportional to their masses and inversely proportional to their squared distance. This translates for travel demand prediction models with the following formulation

$$V_{ij} = k\frac{(A_iA_j)^\alpha}{d_{ij}^\gamma},\ i \neq j,\ d_{ij} = d_{ji}\ , \tag{3.1}$$

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 (3.1) expresses non-directional passenger volume, 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

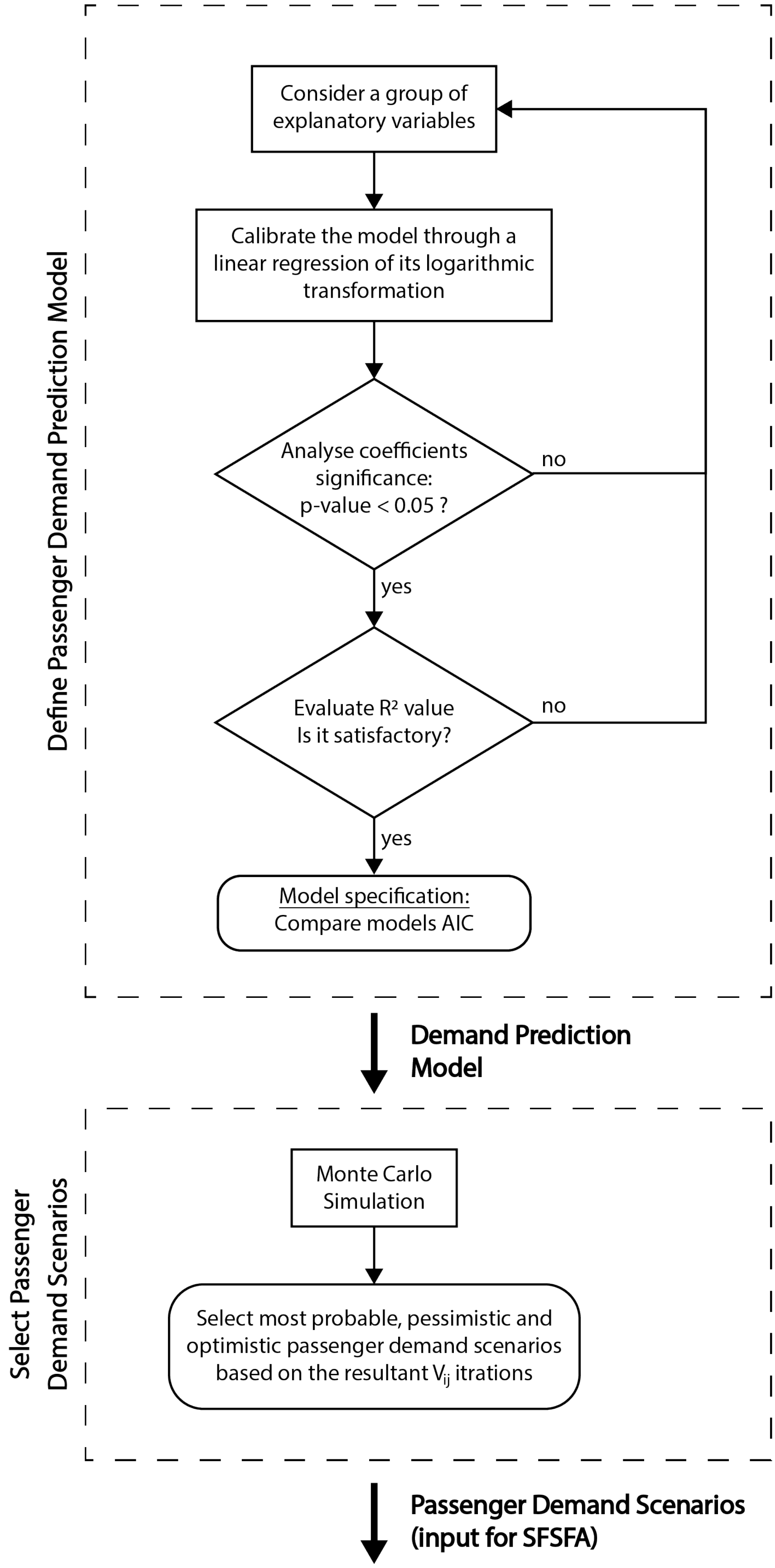


Figure 3.1: Methodological approach for passenger demand estimation diagram.

are the same, thus variables from (3.1) sustain, only allowing different powers α and β for A_i and A_j, respectively [13].

Constant k and parameters α and γ can be estimated by linearizing (3.1) through a logarithmic transformation (base-10 logarithm is used),

$$\log(V_{ij}) = \log(k) + \alpha \log(A_i A_j) - \gamma \log(d_{ij}) \ , \tag{3.2}$$

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.

3.1.2 Explanatory Variables

There are several different approaches to select explanatory variables, as presented in Section 2.1.3. The significance of several explanatory variables suggested in the literature when applied to the specific Rhodes based network case study is analyzed in Section 5.2. The assessment of the importance of the explanatory variables considered is based on the *p-value*. A null-hypothesis is defined as stating there is no relationship between the explanatory variable under discussion and the dependent variable, V_{ij}. The *p-value* describes the probability of the null-hypothesis being true. Therefore, lower *p-values* provide stronger evidence to reject the null-hypothesis and consequently prove that the explanatory variable is significant. The *p-value* threshold to consider that an explanatory variable is proven significant is determined as 0.05, i.e. a 5% significance level will be considered.

The value used to assess the absolute overall quality of the model is the R-squared. However, other concerns like the presence of multicollinearity and overfitting are considered. As Alin [23] stated, multicollinearity refers to the linear relationship between variables, which translates into a lack of orthogonality. In a multivariate linear regression model, an explanatory variable x coefficient can be interpreted as the change in the expected value of the dependent variable y when x is increased by one unit keeping the remaining explanatory variables constant. If the explanatory variables are not independent, this interpretation no longer holds, because a change in one explanatory variable prevents holding other correlated variables constant since they may suffer an associated shift. Therefore, although multicollinearity does not hinder the prediction of new observations if these inferences are made within the range of observed data, it does impair the analysis when one wants to isolate the relationship between each explanatory variable and the dependent variable.

According to Hawkins [24], regression problems (using multiple linear regression, like the present work, or other types of methodology) can be formulated within two distinct primary settings: prediction problems or effect quantification. On the one hand, the objective of this work is to design a model to predict unconstrained demand, fitting in the prediction problem classification. On the other hand, it is useful to gain an understanding of how each predictor affects the dependent variables, because only in that way one can apply a sensibility analysis in respect to each explanatory variable, individually. That is why the correlation between explanatory variables is carefully analyzed in Section 5.2.7.

Overfitting corresponds to not complying with the principle of parsimony, that “calls for using models and procedures that contain all that is necessary for the modeling but nothing more” - Hawkins [24], which leads to models that are too flexible or that include irrelevant components. Adding unnecessary explanatory variables that increase the complexity of the model without a benefit in performance not only is a waste of resources (because one has to gather data for all variables to apply the model) and increases the probability of errors in the database, but also hinders its portability, i.e., makes the model more difficult to reproduce. Irrelevant predictors can also impair the quality of the predictions, as their coefficients impose random variations to the predictions. Thus, the Akaike Information Criterion (AIC) is also calculated, as it is a comparative statistic measure that accounts for both their goodness of fit and their complexity, defined by

$$\text{AIC} = -2l + 2p, \tag{3.3}$$

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

Note that the AIC is not an absolute measure and so it cannot ensure the quality of a model by itself but just compare a model in relation to another model.

3.1.3 Design of Demand Scenarios

The passenger demand prediction model specified through the method described in the previous Section 3.1.2 is used to obtain three different passenger demand scenarios to use as input for the Route and Flight Schedule Optimization program described in Section 3.2. The present section firstly describes the method through which the number of passengers estimates are obtained and then explains their translation into the three input scenarios.

Monte Carlo Simulation

Once a suitable set of explanatory variables is identified and their respective coefficients are estimated, the model obtained is ready to use on the estimates of passenger volumes, even on previously non-existent connections. A Monte Carlo (MC) simulation, schematized in Figure 3.2, 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 (3.2). The sampled coefficients are applied to the model to compute V_{ij} for a wanted number of connections c, between airports i and j. Each V_{ij} observation is stored as a vector V of length c.

There is extensive literature on MC stopping Rules. Mendo and Hernando [25] identify two methods to determine sample size in MC simulations: fixing the sample size c beforehand based on *a priori* knowledge, or using a sequential stopping procedure, where the sample size c is determined according

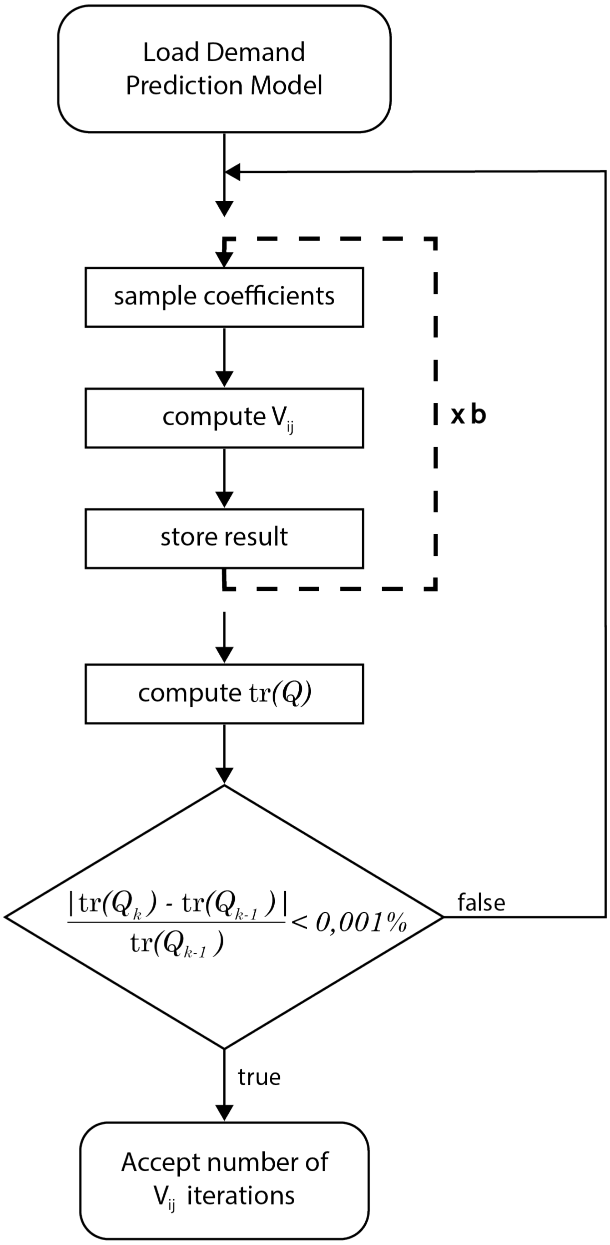


Figure 3.2: Monte Carlo simulation diagram.

to the outcome of the simulation itself. The latter method was chosen and the criterion to stop iterating is based on the convergence of the trace of the sample covariance 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

$$Q = \frac{1}{N-1}\sum_{n=1}^{N}(V_n - \overline{V})(V_n - \overline{V})^T \quad , \tag{3.4}$$

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.

Demand Scenarios Selection

The distribution of the resultant MC iterations is then used to define the most probable, a pessimistic, and an optimistic scenario, using the median and percentiles. Making use of the distribution of each route passenger volume V_{ij} to define their own most probable, pessimistic, and optimistic scenarios would disregard the covariance between the different routes. Consequently, overlaying every route pessimistic/optimistic scenario would result in exaggerated overall pessimistic/optimistic scenarios.

To avoid this problem, the distribution of the sum of passengers for every route can be used instead. Figure 3.3 allows an easier understanding of the method described hereafter. $P_{V_{ij}}$ 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 passengers from all routes.

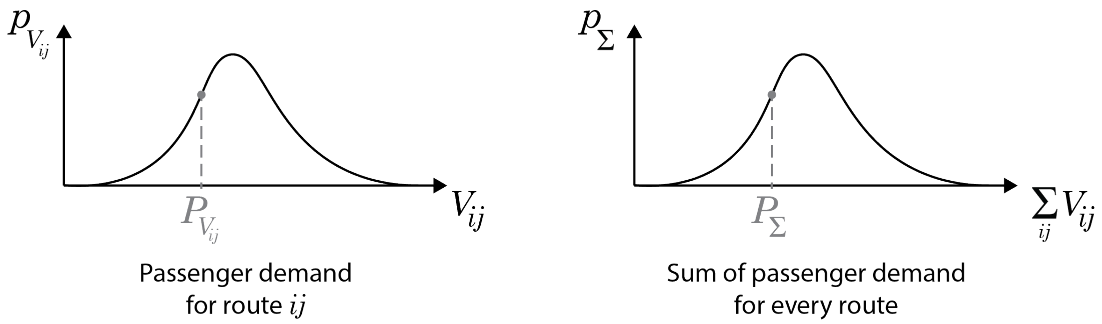


Figure 3.3: Illustrative diagram of the distribution of iterations regarding passenger demand for route ij and the sum of passenger demand for every route, respectively.

This approach raises a new question about the choice of the scenarios, since 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). To address this issue, 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

$$\sum_{ij} (P_{V_{ij}} - V_{ij_m})^2 \tag{3.5}$$

is selected as the scenario referring to the said percentile.

For example, to compute the overall 50% percentile, there may be $M = 6$ iterations with that same total number of passengers $P_{\sum}$; for each of those 6 competing iterations, the absolute difference between each route iterated number of passengers and their individual 50% percentile (according to their own Vij distribution) is computed; the iteration m presenting the lowest sum of squared differences of all routes among those 6 is selected.

3.2 Route and Flight Schedule Optimization

As shown in Section 2.2, there is some literature aiming to develop fully integrated FSFA models. Pita et al. [5] designs a model for a subsidized network context considering:

- airline's costs: flight costs (as a function of traveled distance and aircraft type); and off-base costs (consisting on charges whenever an aircraft is not parked at the home base);
- airport's costs: operational costs, including employee wages, operating equipment, energy, supplies, materials, and outsourcing expenses (as a function of the opening times of an airport); and non-aero-nautical revenues, including commercial, retail, and parking revenues (as a function of passenger volume);
- passenger's costs: on-board time costs; waiting on ground time costs, for multiple stops itineraries; and schedule delay costs (accounting for the difference between the scheduled time for a flight and the passenger's desired time to travel that flight).

Leandro et al. [6] takes a more simplistic approach, where the only costs considered are: the costs for the airline related to the operation of flights and to parking for longer than 2 hours, and the social costs related to passengers' time onboard and waiting on ground between flight legs.

As mentioned before, the objective of this work is to obtain a robust decision support framework, which considers the uncertainty inherent to the passenger demand estimation due to unpredictable passenger behavior and unexpected exogenous events. Thus, the contribution of the developed optimization model is the introduction of pessimistic and optimistic scenarios of passenger demand. Therefore, the focus of the developed model is not to enhance state-of-the-art models by considering more detailed costs in order to get closer to a *real-world* application, but to explore robustness. With this in mind, a simpler approach based on Leandro et al. [6] was followed to be then further extended by assuming a robust optimization perspective and carefully designing different passenger demand scenarios to accomplish that.

3.2.1 Problem Formulation

The formulation of the Mixed Integer Linear Programming (MILP) problem, to be solved with FICO Xpress software, includes the definition of the sets, parameters, decision variables, and constraints, presented in this section.

Hereafter, every pair of two different airports designated as a *link* corresponds to the origin and destination desired by the passenger. To satisfy the demand for a link, 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.

To accommodate the clustering problem relaxation in Section 6.2.2, the permitted direct routes and multiple stops itineraries are pre-computed and fed directly to the optimization software. Therefore multiple stops itineraries are not defined as a combination of airports or direct routes, like in Leandro et al. [6], but sets of direct routes, one, two, and three stops itineraries are defined *a priori* instead.

Sets

$A \quad = \{1, ..., NA\}$: set of airports within the network

$L \quad = \{1, ..., NL\}$: set of links (all combinations between two airports)

$F \quad = \{1, ..., NF\}$: set of direct routes

$G \quad = \{1, ..., NG\}$: set of one stop itineraries

$H \quad = \{1, ..., NH\}$: set of two stop itineraries

$I \quad = \{1, ..., NI\}$: set of three stop itineraries

$R \quad = \{1, ..., NR\}$: set of aircraft types available

$T \quad = \{1, ..., NT\}$: set of time steps in an operational day

$D \quad = \{1, ..., ND\}$: set of operational days in a week

$W \quad = \{1, ..., NW\}$: set of permitted waiting times (expressed in number of time steps)

$S \quad = \{1, ..., NS\}$: set of passenger demand scenarios to consider

Parameters

$z(r)$: number of aircraft of each type r, with $r \in R$

$s(r)$: seat capacity of each aircraft type r, with $r \in R$

$x_{min}(f)$: minimum number of flights on link l per week, with $l \in L$

$s_{min}(f)$: minimum number of seats for link l per week, with $l \in L$

$a_l(l)$: array with departure and arrival airports of link l, with $l \in L$

$a_f(f)$: array with departure and arrival airports of route f, with $f \in F$

$a_g(g)$: array with departure, stop, and arrival airports of itinerary g, with $g \in G$

$a_h(h)$: array with departure, stops, and arrival airports of itinerary h, with $h \in H$

$a_i(i)$: array with departure, stops, and arrival airports of itinerary i, with $i \in I$

$t_f(f)$: travel time of route f, with $f \in F$

$t_g(g)$: travel time (onboard) of itinerary g, with $g \in G$

$t_h(h)$: travel time (onboard) of itinerary h, with $h \in H$

$t_i(i)$: travel time (onboard) of itinerary i, with $i \in I$

$q(l, s)$: passenger demand for link l in scenario s, with $l \in L$ and $s \in S$

$prob(s)$: probability of scenario s, with $s \in S$

$c_F(r)$: cost of flying the airplane type r, per time period, with $r \in R$

$c_A(a, r)$: cost of having an aircraft of type r on the ground for one time step, in airport a, with $r \in R$ and $a \in A$

c_B : time cost of being onboard, for a passenger, per time period

c_G : time cost of being waiting on ground, for a passenger, per time period

All parameters have to be defined as inputs to the program. The constants to define the number of elements of sets, NA, NF, NG, NH, NI, NR, NT, ND, NW, and NS, are needed inputs as well, except for NL, that can be derived as $NL = NA \times (NA - 1)$.

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, with $f \in F$, $r \in R$, $t \in T$, and $d \in D$

$y(a, t, d, r)$: number of aircraft of type r on ground at airport a during time step t of day d, with $r \in R$, $a \in A$, $t \in T$, and $d \in D$

$u_D(f, t, d, s)$: number of passengers placed on direct route f, departing at time t of day d, considering scenario s, with $f \in F$, $t \in T$, $d \in D$, and $s \in S$

$u_1(g, t, d, w, s)$: 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, with $g \in G$, $t \in T$, $d \in D$, $w \in W$, and $s \in S$

$u_2(h, t, d, w_1, w_2, s)$: 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, with $h \in H$, $t \in T$, $d \in D$, $w_1, w_2 \in W$, and $s \in S$

$u_3(i, t, d, w_1, w_2, w_3, s)$: 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, with $i \in I$, $t \in T$, $d \in D$, $w_1, w_2, w_3 \in W$, and $s \in S$

$gc(a, t, d, r)$: 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, with $r \in R$, $a \in A$, $t \in T$, and $d \in D$, and takes 0 otherwise

Figure 3.4 shows a graphical representation of an example, aiming to ease the notation interpretation. As depicted, $a_f(f_1) = [5, 4]$ and $a_f(f_2) = [4, 6]$ are legs of one stop itinerary g. Knowing that $t_f(f_1) = 2$ and $t_f(f_2) = 3$, g can be fully characterized by $a_g(g) = [5, 4, 6]$ and $t_g(g) = 5$. Assuming 10 passengers were allocated to fly g to depart in instant $t = 2$ and with stop waiting time $w = 1$, $u_1(g, 2, d, 1, s) = 10$, in day d for scenario s. Considering just one aircraft, this would imply that that aircraft of type r verified $x(f_1, 2, d, r) = 1$, $y(4, 4, d, r) = 1$, and $x(f_2, 5, d, r) = 1$. Note that the definition of one stop itinerary g does not depend on the stop waiting time, as the muted itinerary representation only differs by starting on $t = 4$ and having $w = 2$, and still corresponds to route g, but carrying $u_1(g, 4, d, 2, s)$ passengers.

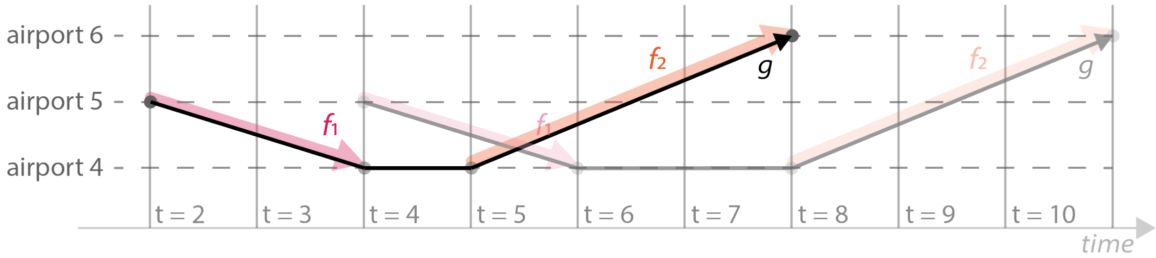


Figure 3.4: Graphical representation of one stop itinerary example, g.

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

$$\sum_{a \in A} y(a, t, d, r) + \sum_{\substack{f \in F, t' \in T \\ t' < t < t' + t_f(f)}} x(f, t', d, r) = z(r), \forall\, t \in T, d \in D, r \in R \tag{3.6}$$

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

$$\begin{gathered} y(a, t-1, d, r) + \sum_{\substack{f \in F \\ a_f(f,2) = a}} x(f, t - t_f(f), d, r) = y(a, t, d, r) + \sum_{\substack{f \in F \\ a_f(f,1) = a}} x(f, t, d, r), \\ \forall\, a \in A, t \in T \backslash \{1\}, d \in D, r \in R \end{gathered} \tag{3.7}$$

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

$$y(NA, 1, d, r) + \sum_{\substack{f \in F \\ a_f(f,1) = NA}} x(f, 1, d, r) = z(r), \forall\, d \in D, r \in R \tag{3.8}$$

$$y(NA, NT, d, r) + \sum_{\substack{f \in F \\ a_f(f,2) = NA}} x(f, NT + 1 - t_f(f), d, r) = z(r), \forall\, d \in D, r \in R \tag{3.9}$$

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, as schematized in Figure 3.5, were aircraft of type r flying $x(f,5,d,r)$ carries passengers $u_2(h,t-w_1-t_f(f_1),d,w_1,w_2,s)$, i.e. $u_2(h,2,d,1,1,s)$ in day d of scenario s

$$\begin{aligned}
&u_D(f,t,d,s) + \sum_{\substack{g\in G,w\in W\\ a_g(g,1)=a_f(f,1)\wedge a_g(g,2)=a_f(f,2)}} u_1(g,t,d,w,s)\\
&+ \sum_{\substack{g\in G,f_1\in F,w\in W\\ a_g(g,1)=a_f(f_1,1)\wedge a_g(g,2)=a_f(f_1,2)\\ a_g(g,2)=a_f(f,1)\wedge a_g(g,3)=a_f(f,2)}} u_1(g,t-w-t_f(f_1),d,w,s)\\
&+ \sum_{\substack{h\in H,w_1,w_2\in W\\ a_h(h,1)=a_f(f,1)\wedge a_h(h,2)=a_f(f,2)}} u_2(h,t,d,w_1,w_2,s)\\
&+ \sum_{\substack{h\in H,f_1\in F,w_1,w_2\in W\\ a_h(h,1)=a_f(f_1,1)\wedge a_h(h,2)=a_f(f_1,2)\\ a_h(h,2)=a_f(f,1)\wedge a_h(h,3)=a_f(f,2)}} u_2(h,t-w_1-t_f(f_1),d,w_1,w_2,s)\\
&+ \sum_{\substack{h\in H,f_1,f_2\in F,w_1,w_2\in W\\ a_h(h,1)=a_f(f_1,1)\wedge a_h(h,2)=a_f(f_1,2)\\ a_h(h,2)=a_f(f_2,1)\wedge a_h(h,3)=a_f(f_2,2)\\ a_h(h,3)=a_f(f,1)\wedge a_h(h,4)=a_f(f,2)}} u_2(h,t-w_2-t_f(f_2)-w_1-t_f(f_1),d,w_1,w_2,s)\\
&+ \sum_{\substack{i\in I,w1,w_2,w_3\in W\\ a_i(i,1)=a_f(f,1)\wedge a_i(i,2)=a_f(f,2)}} u_3(i,t,d,w_1,w_2,w_3,s)\\
&+ \sum_{\substack{i\in I,f_1\in F,w_1,w_2,w_3\in W\\ a_i(i,1)=a_f(f_1,1)\wedge a_i(i,2)=a_f(f_1,2)\\ a_i(i,2)=a_f(f,1)\wedge a_i(i,3)=a_f(f,2)}} u_3(i,t-w_1-t_f(f_1),d,w_1,w_2,w_3,s)\\
&+ \sum_{\substack{i\in I,f_1,f_2\in F,w_1,w_2,w_3\in W\\ a_i(i,1)=a_f(f_1,1)\wedge a_i(i,2)=a_f(f_1,2)\\ a_i(i,2)=a_f(f_2,1)\wedge a_i(i,3)=a_f(f_2,2)\\ a_i(i,3)=a_f(f,1)\wedge a_i(i,4)=a_f(f,2)}} u_3(i,t-w_2-t_f(f_2)-w_1-t_f(f_1),d,w_1,w_2,w_3,s)\\
&+ \sum_{\substack{i\in I,f_1,f_2,f_3\in F,w_1,w_2,w_3\in W\\ a_i(i,1)=a_f(f_1,1)\wedge a_i(i,2)=a_f(f_1,2)\\ a_i(i,2)=a_f(f_2,1)\wedge a_i(i,3)=a_f(f_2,2)\\ a_i(i,3)=a_f(f_3,1)\wedge a_i(i,4)=a_f(f_3,2)\\ a_i(i,4)=a_f(f,1)\wedge a_i(i,5)=a_f(f,2)}} u_3(i,t-w_3-t_f(f_3)-w_2-t_f(f_2)-w_1-t_f(f_1),d,w_1,w_2,w_3,s)\\
&= \sum_{r\in R} s(r)x(f,t,d,r), \forall f\in F, t\in T, d\in D, s\in S
\end{aligned} \tag{3.10}$$

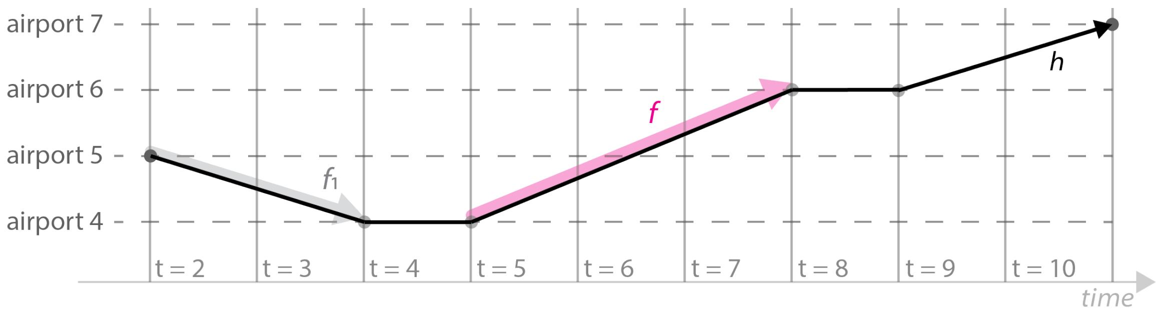


Figure 3.5: Graphical representation of two stop itinerary example, h.

Demand satisfaction : for every scenario, s, the demand between every link l corresponds to the number of passengers carried either in direct routes or up to three stops itineraries

$$\begin{aligned} q(l,s) = & \sum_{\substack{f \in F, t \in T, d \in D \\ a_f(1)=a_l(1) \wedge a_f(2)=a_l(2)}} u_D(f,t,d,s) \\ & + \sum_{\substack{g \in G, t \in T, d \in D, w \in W \\ a_g(1)=a_l(1) \wedge a_g(3)=a_l(2)}} u_1(g,t,d,w,s) \\ & + \sum_{\substack{h \in H, t \in T, d \in D, w1, w2 \in W \\ a_h(1)=a_l(1) \wedge a_h(4)=a_l(2)}} u_2(h,t,d,w_1,w_2,s) \\ & + \sum_{\substack{i \in I, t \in T, d \in D, w1, w2, w3 \in W \\ a_i(1)=a_l(1) \wedge a_i(5)=a_l(2)}} u_3(i,t,d,w_1,w_2,w_3,s), \forall\, l \in L, s \in S \end{aligned} \tag{3.11}$$

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*

$$\sum_{t \in T} \sum_{d \in D} \sum_{r \in R} x(f,t,d,r) \geq xmin(f), \forall\, f \in F \tag{3.12}$$

$$\sum_{t \in T} \sum_{d \in D} \sum_{r \in R} s(r) x(f,t,d,r) \geq smin(f), \forall\, f \in F \tag{3.13}$$

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)

$$\begin{gathered} gc(a,t,d,r) \geq y(a,t,d,r) + y(a,t+1,d,r) + y(a,t+2,d,r) + y(a,t+3,d,r) - 3.5, \\ \forall\, a \in A, t \in T \backslash \{NT-1, NT-2, NT-3\}, d \in D, r \in R \end{gathered} \tag{3.14}$$

Objective Function

The minimization problem follows

$$\text{minimize} \sum_{k=1}^{9} O_k \quad , \tag{3.15}$$

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

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

$$O_1 = \sum_{f \in F} \sum_{t \in T} \sum_{d \in D} \sum_{r \in R} c_F(r)\; t_f(f)\; x(f,t,d,r) \tag{3.16}$$

Parking aircraft costs: 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

$$O_2 = \sum_{a \in A} \sum_{t \in T} \sum_{d \in D} \sum_{r \in R} c_S(r)\ gc(a,t,d,r) \tag{3.17}$$

Passengers onboard time costs: sum of passengers onboard time cost for direct routes f and up to three stops itineraries g, h, i

$$O_3 = \sum_{f \in F} \sum_{t \in T} \sum_{d \in D} \sum_{s \in S} c_B(r)\ t_f(f)\ u_D(f,t,d,s)\ p(s) \tag{3.18}$$

$$O_4 = \sum_{g \in G} \sum_{t \in T} \sum_{d \in D} \sum_{s \in S} \sum_{w \in W} c_B(r)\ t_g(g)\ u_1(g,t,d,w,s)\ p(s) \tag{3.19}$$

$$O_5 = \sum_{h \in H} \sum_{t \in T} \sum_{d \in D} \sum_{s \in S} \sum_{w_1 \in W} \sum_{w_2 \in W} c_B(r)\ t_h(h)\ u_2(h,t,d,w_1,w_2,s)\ p(s) \tag{3.20}$$

$$O_6 = \sum_{i \in I} \sum_{t \in T} \sum_{d \in D} \sum_{s \in S} \sum_{w_1 \in W} \sum_{w_2 \in W} \sum_{w_3 \in W} c_B(r)\ t_i(i)\ u_3(i,t,d,w_1,w_2,w_3,s)\ p(s) \tag{3.21}$$

Passengers waiting time costs: sum of passengers waiting on ground between legs time cost for up to three stops itineraries g, h, i

$$O_7 = \sum_{g \in G} \sum_{t \in T} \sum_{d \in D} \sum_{s \in S} \sum_{w \in W} c_W(r)\ w\ u_1(g,t,d,w,s)\ p(s) \tag{3.22}$$

$$O_8 = \sum_{h \in H} \sum_{t \in T} \sum_{d \in D} \sum_{s \in S} \sum_{w_1 \in W} \sum_{w_2 \in W} c_W(r)\ (w_1 + w_2)\ u_2(h,t,d,w_1,w_2,s)\ p(s) \tag{3.23}$$

$$O_9 = \sum_{i \in I} \sum_{t \in T} \sum_{d \in D} \sum_{s \in S} \sum_{w_1 \in W} \sum_{w_2 \in W} \sum_{w_3 \in W} c_W(r)\ (w_1 + w_2 + w_3)\ u_3(i,t,d,w_1,w_2,w_3,s)\ p(c) \tag{3.24}$$

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. When direct, one, two, or three stop itineraries depart on the beginning of the day not from the hub, arrive at the end of the day not to the hub, or do not arrive before the end of the day at all, the respective cells of u_D, u_1, u_2, or u_3 are not created. 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.

3.2.2 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, different relaxation approaches were explored.

Firstly, the number of possible itineraries was reduced, based on a cluster strategy, and fed directly to the optimization problem. The airports were grouped in clusters according to the proximity to each other. Thereafter entry and exit airports for each cluster were defined and restrictions on inter-cluster traveling were applied. An important aspect to take into account is that applying such restrictions limits the possibility to travel between airports from distinct clusters, possibly forcing non-direct itineraries to have more stops.

Secondly, the demand was divided in parts that would be ran separately, significantly reducing the number of time steps to consider, and therefore the computational effort. This is a delicate problem relaxation, as it could potentially lead to a problem not as close to the original one as it is desired.

These two approaches are detailed in Section 6.2.2, upon the application to the Greek islands case study in particular.

Chapter 4

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 them inhabited, that will not be accounted for. Figure 4.1 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, presented in Table 4.1.

Table 4.1: All discussed airports and respective IATA codes.

Island	Airport	IATA code
-	Athens International Airport	ATH
Karpathos	Karpathos National Airport	AOK
Crete (Chania)	International Airport “Ioannis Daskalogiannis”	CHQ
Crete (Heraklion)	Heraklion Airport “N. Kazantzakis”	HER
Chios	Chios Airport “Omiro”	JKH
Kalymnos	Kalymnos National Airport	JKL
Astypalaia	Astypalaia National Airport	JTY
Kos	Kos Airport “Ippokrati”	KGS
Kasos	Kasos Municipal Airport	KSJ
Kastelorizo	Kastelorizo Municipal Airport	KZS
Leros	Leros Municipal Airport	LRS
Limnos	Limnos Airport “Ifestos”	LXS
Lesvos	Mytilini Airport “Od. Elytis”	MJT
-	Thessaloniki Airport Makedonia	SKG
Samos	Samos Airport “Aristarchos of Samos”	SMI
Rhodes	Rhodes Airport “Diagoras”	RHO

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). A PSO Inventory Table as of 18/9/2019 is publicly

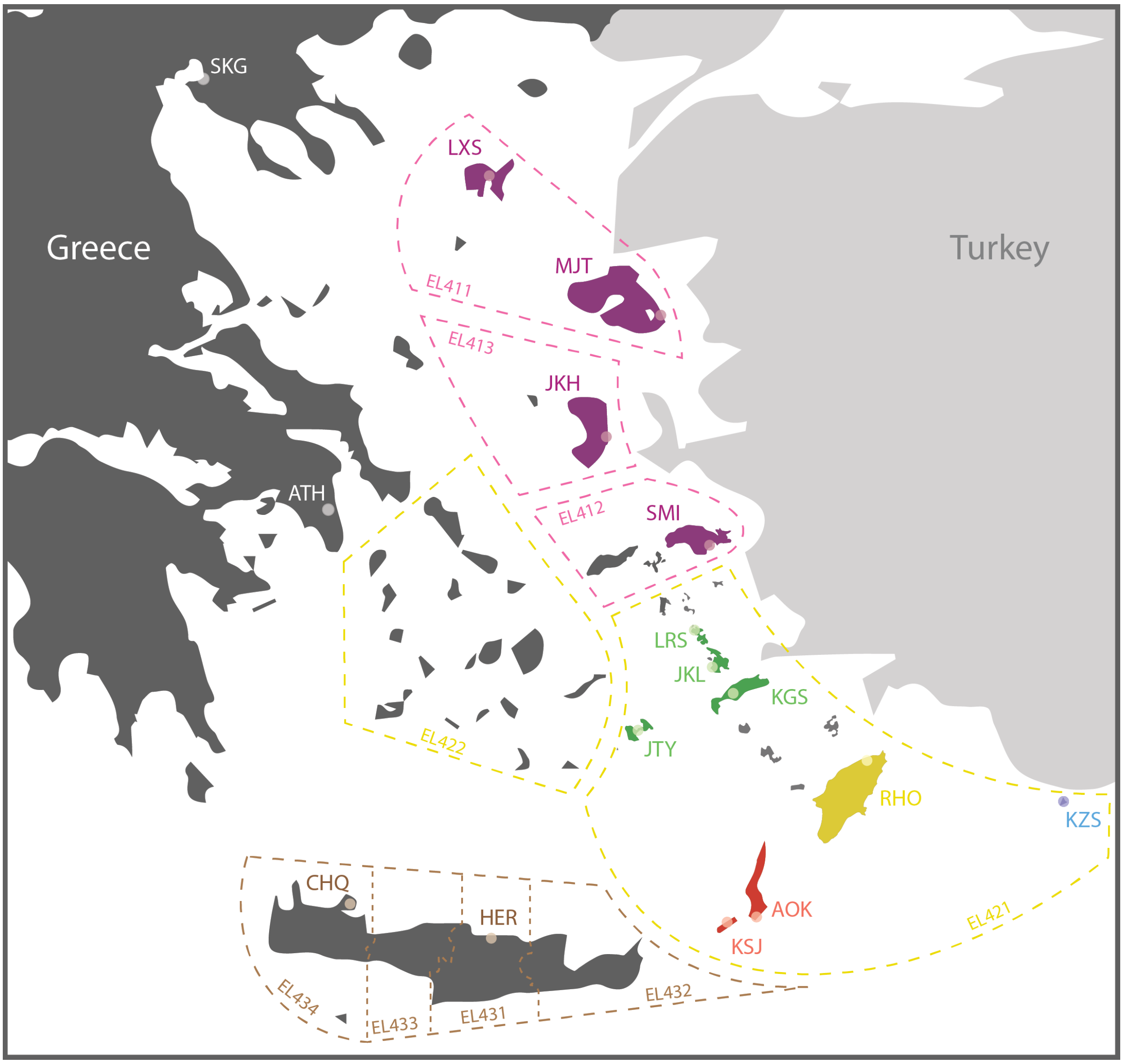


Figure 4.1: Case Study Greek Islands map.

available at the European Commission Website [26]. This list presents the four groups of PSOs including RHO, shown in Table 4.2. In addition to the airports served, tender requirements, and some operational information on these routes is also available. Only the relevant information to understand the present work is presented.

All four entries of Table 4.2 are under a restricted PSO regime. The three first columns stand for number of months of operation per year, minimum number of annual seats required, and weekly frequencies required, respectively. The three values on weekly flight frequency column stand for low, medium, and high demand seasons, respectively (e.g. [3/4/6] stands for a minimum of 3 flights per week on low demand season, 4 on medium demand season, and 6 on high demand season).

All airports mentioned in Table 4.2 are highlighted in Figure 4.1 with a color code to ease routes distinction by PSO group (the "PSO group" nomenclature will be used accordingly hereafter). Rhodes Diagoras Airport is colored yellow, as it is the common airport for all four analyzed routes (RHO will

Table 4.2: PSO with connections to Rhodes Diagoras Airport.

Airports	Months /year	Min. seats /year	Min. flights /week	Annual compensation (2018)	Aircraft type
LXS - MJT - JKH - SMI - RHO	12	6050	2/2/3	€890.009,75	BAE JS41 (29 seats)
RHO - AOK - KSJ	12	13200	3/4/6	€795.000,00	ATR 42-320 (48 seats)
RHO - KZS	12	800	3/4/6	€919.199,00	Dash 8-100 (37 seats)
RHO - KGS - JKL - LRS - JTY	12	6050	2/2/3	€1.089.000,00	ATR 42-320 (48 seats)

be considered a “PSO group” by itself according to this nomenclature). The purple group connects 4 North Aegean (*Voreio Aigaio*) islands with Rhodes. All the remaining colored islands belong to the Dodecanese archipelago, which is part of the South Aegean (*Notio Aigaio*). The network to be optimized, by applying the method developed in the present work consists of these three last groups, as a 8 islands confer an adequate complexity to the case study.

NUTS classification is a system created by the European Union to divide territory for regional statistics collection and analysis, among other purposes. There are NUTS levels 1, 2, and 3, where 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.

North Aegean and South Aegean correspond to NUTS 2 regions and are identified in the map by the pink and yellow dashed contours, respectively. NUTS 3 regions are also marked in Figure 4.1, as it is important to acknowledge that many of the islands under analysis belong to the Dodecanese NUTS 3 region, with code EL421. 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. As for the quality of transport infrastructure, according to a briefing requested by the Committee on Transport and Tourism (TRAN) of the European Parliament in 2018 [27], *”the Greek ports (...) are rated relatively low, however the quality of (...) air transports are around the EU average”*. The current ferry transportation service presents a number of deficiencies, such as low frequencies, high travel times, and poor quality of services: during the Summer passenger demand peak, vessels get overcrowded and ship departures and arrivals get delayed; during the off-peak periods, travel frequency is reduced and fast boats are retracted from service, because of their higher cost and the difficulty to sail under winter weather [19].

Tourism is an important contributor to the Greek economy, as it directly contributed 6.4% Gross Value Added (GVA) to the Greece’s GDP while supporting nearly 366 thousand jobs, approximately 10% of jobs in Greece, as of 2016 [27]. 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 [28]. Besides being a beach holiday destination, Greece has 18 properties inscribed on the UNESCO World Heritage Sites List, as of 2021, 5 of them located in Aegean islands [29]. South Aegean islands 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 [30], 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 [27].

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 [31].

Chapter 5

Demand Prediction Model

This chapter presents the implementation of the predictive methods presented in Section 3.1 to the Rhodes based PSO network case study. Firstly, the available data is explored and, based on that, the reasoning behind the choice of gravity models as the most suitable type of prediction model. Section 5.1 further describes how the data was processed and used to design the prediction models. Section 5.2 presents the analysis of several models with different combinations of explanatory variables. Section 5.3 describes how the different scenarios to feed the Route and Flight Schedule Optimization program are obtained.

5.1 Description of available data-set

The airport infrastructure manager Fraport Greece provided detailed information on all domestic arrivals and departures on Rhodes Diagoras airport for August and November, years 2018, 2019, and 2020, including each flight time of arrival/departure to/from RHO, number of passengers, load factor, and type of aircraft operating it, depicted in Table 5.1. Statistical data on all Greek Airports, with the monthly number of domestic passenger arrivals (Arr) and departures (Dep), is publicly available in the Hellenic Republic Aviation Authority website [32]. Econometric data was mainly taken from the Hellenic Statistical Authority, namely from the 2011 Census. Note that a fair amount of the analyzed islands belongs to the same NUTS 3 region, limiting the usability of such data, as shown in Section 5.2.2.

Table 5.1: Fraport data depiction.

	Year	Month	Day	Dep.	Arr.	Time	Pax.	Load factor	Aircraft type
ex.:	2019	Aug.	22	RHO	HER	16:15	48	100%	ATR 42-320
	2020	Nov.	8	KZS	RHO	09:50	1	2,7%	Dash 8-100

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 the 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, Fraport Greece data 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. Thessaloniki and Athens connections were disregarded, as they are not islands and, therefore, are not within the intended model scope. Flights from Chania International Airport ”Ioannis Daskalogiannis” (CHQ) were summed to Heraklion Airport ”N. Kazantzakis” (HER), as they are in the same island. Thereby, valid entries for V_{ij} sum 82 valid data points, corresponding to the sum of passengers flying a link, throughout a month, as depicted in Table 5.2 (for example, there were ten flights from KZS to RHO in November 2020, then the resultant entry corresponds to the sum of passengers on those flights).

Table 5.2: Data entries exported from Fraport data depiction.

	Year	Month	Dep. (i)	Arr. (j)	Pax. (V_{ij})
ex.:	2019	Aug.	RHO	HER	2288
	2020	Nov.	KZS	RHO	59

Figure 5.1 presents the distribution of passengers per year and per season according to the 82 data points considered. One can clearly observe the effect of seasonality and the decrease in 2020 due to the Covid-19 pandemic.

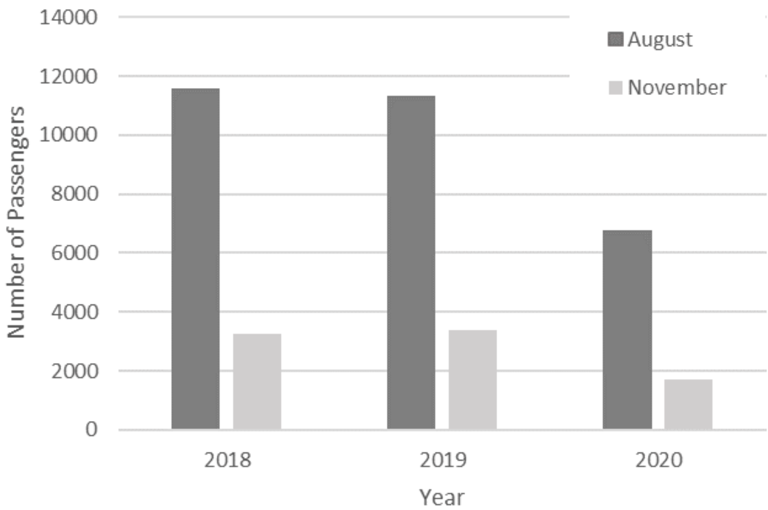


Figure 5.1: Domestic Passenger Volume in Rhodes Diagoras Airport, (except ATH and SKG, i.e. regarding the 82 data points considered).

There are also great disparities in passenger demand volume among the links under study, as one can verify in Table 5.3, that presents the main descriptive statistics of the sample. The table includes every explanatory variable used in the models presented hereafter, with d for distance between airports, P for Population, G for Per Capita GDP, and T for number of nights spent by tourists in that island.

Table 5.3: Data-set descriptive statistics.

Statistics	V_{ij}	Dep_i	Arr_j	d_{ij} (km)	P_i	P_j	G_i (€)	G_j (€)	T_i	T_j
Mean	463	20597	20319	206	107657	109332	15170	15221	230	235
Median	123	19805	24317	154	115000	115000	15838	15838	296	296
Std. Dev.	836	18361	18361	112	154206	153770	1449	1396	129	127
Minimum	4	22	15	100	492	492	11471	11471	15	23
Maximum	4090	68639	75783	465	623065	623065	15838	15838	562	562

It is important to keep in mind that this data is being used as a proxy for unconstrained demand, but it is clearly influenced by the offered service. However, the interactions between supply and demand in network design (Birolini et al. [33]) are out of the scope of the present work. Additionally, the fact that many passengers are carried in multiple stops itineraries impairs the interpretation of this data, as part of the passengers carried on route f, that departs from i and arrives to j, can be just flying f as a leg of an itinerary with origin different than i and/or destination different than j, as explained later in Section 5.3.

5.2 Identification of Explanatory Variables

None of the models proposed in the literature from Section 2.1 presented the most suitable set of explanatory variables for this particular case study, as it is shown in this section. Several different new combinations of those factors, taking into account the particularities of the Greek islands case study, were explored and some are presented hereafter. The reason for considering those variables in the first place and their drawbacks are discussed, as well as an analysis of their significance. MATLAB was the software used to estimate the linear regression models through a standard least-squares method.

Section 5.2.7 presents a final analysis with the comparison of the most suitable models.

5.2.1 *Model 1*: Service-related variables

The first set of variables considered included the sum of passenger arrivals and departures for each airport as the only attractiveness factor, as suggested by Doganis [14]. *Model 1* corresponds to a similar version, with the number of passenger departures, Dep_i, and arrivals, Arr_j, providing direction to the model, and is described as:

$$V_{ij} = k\, d_{ij}^{\gamma}\, Dep_i^{\alpha}\, Arr_j^{\beta}\ , \tag{5.1}$$

with the geographical distance d_{ij} between airports expressed in kilometers. The constant k and explanatory variables coefficients γ, α, and β, for d_{ij}, Dep_i, and Arr_j, respectively, are estimated by linearizing the gravity model equation (5.1) through a logarithmic transformation and performing a linear regression using the available data, as described in Section 3.1.1.

The results of the linear regression are presented in Table 5.4. The effect of the distance between airports d_{ij} is not significant, with a p-value of 0.20, whilst the passenger flow departing and arriving at the origin and destination airports, Dep_i and Arr_j, have significant and similar p-values of 0.00019 and 0.00012, respectively, which shows evidence of their relation with the dependent variable. Note that coefficients α and β are positive, as a higher airport passenger flow is associated with a higher passenger volume for the connection between them, as expected, and both assume the same value of 0.26, backing the non-directional approach taken by Doganis [14]. However, the R-squared value of 0.26 is low and not acceptable.

It was known that *Model 1*, as it is, would distort attraction of the airports by magnifying the importance of airports that serve as hub or as stop of a non-direct itinerary. When adding other explanatory variables to a model with these service-related variables, Dep_i and Arr_j, the added variables (d_{ij}, for example) present low levels of significance, which could be explained by an excessive correspondence between islands airports' passenger flow (Dep_i and Arr_j) and the passenger volume of their connections with Rhodes (V_{ij} used for calibration). Since the objective of this model is to predict unconstrained demand, the approach to rely on service-related explanatory variables is not considered the most suitable.

5.2.2 *Model 2*: Geo-economic variables

Model 2 is based on geo-economic explanatory variables rather than the service-related variables used in *Model 1*. This approach is inspired by Grosche et al. [13], by including Population P_i of island i (according to 2011 Census) and Per Capita GDP G_i (according to NUTS 3 data for 2018 to which island i belongs to):

$$V_{ij} = k\, d_{ij}^{\gamma}\, P_i^{\delta}\, P_j^{\epsilon}\, G_i^{\zeta}\, G_j^{\eta}\ . \tag{5.2}$$

The results of the linear regression of the logarithmic transformation of (5.2) are presented in Table 5.4. The population coefficients δ and ϵ assume a positive value with a high level of significance, confirming the positive effect on an island's attractiveness for it to be more populated. As expected, the Per Capita GDP was not deemed significant (p-values of both ζ and η for variables G_i and G_j, respectively, were higher than 0.05), as the variable admits the same value for islands from the same NUTS 3 region (for example, AOK, JTY, KGS, KSJ, KZS, and RHO take the same value). Distance d_{ij} also shows a low level of significance and the R-squared value of the model is low.

5.2.3 *Model 3*: Tourism and seasonality

Other factors that greatly influence air transport passenger volume, particularly on this case study, are seasonality and tourism, as pointed by Sambracos and Rigas [16]. Having this in mind, *Model 3* includes 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:

$$V_{ij} = k\, d_{ij}^{\gamma}\, P_i^{\delta}\, P_j^{\epsilon}\, T_i^{\theta}\, T_j^{\iota}\, S^{\lambda}\ . \tag{5.3}$$

Note that dummy variable S cannot take the common values of 0 and 1 because of the nature of the Gravity model. It is not possible for S to admit value 0, as it would be impossible to perform the logarithmic transformation. When $S = 1$ (low season), it is equivalent to not affecting the prediction, when $S = 10$ (high season), the prediction will be multiplied by the factor 10^{λ}, as the base-10 logarithm is used for the transformation.

The results of the linear regression of the logarithmic transformation of (5.3) are presented in Table 5.4. All added variables were shown to be significant, with p-values lower than 0.05. 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}.

According to the fundamental idea of gravity models, attractiveness between islands should be brought down by higher distances, leading to $\gamma < 0$. However, 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, which may reduce air travel passenger volume V_{ij}. This could explain why the distance coefficient γ, with a good level of significance, assumed a positive value.

The R-squared value of 0.4 for *Model 3* is much higher than the models presented previously and the AIC value is lower.

Table 5.4: *Models 1*, *2*, and *3* coefficient estimates and p-values.

		Model 1 (5.1)		*Model 2* (5.2)		*Model 3* (5.3)	
Coef.	Var.	Estimate	p-value	Estimate	p-value	Estimate	p-value
$\log(k)$	-	0.14	0.85	-31	0.04	-6.8	<0.01
γ	d_{ij}	0.13	0.70	0.35	0.47	1.5	<0.01
α	Dep_i	0.23	<0.01	-	-	-	-
β	Arr_j	0.23	<0.01	-	-	-	-
δ	P_i	-	-	0.35	<0.01	0.17	0.03
ϵ	P_j	-	-	0.36	<0.01	0.19	0.01
ζ	G_i	-	-	3.8	0.05	-	-
η	G_j	-	-	3.2	0.10	-	-
θ	T_i	-	-	-	-	0.91	<0.01
ι	T_j	-	-	-	-	0.79	<0.01
λ	S	-	-	-	-	0.38	<0.01
R^2		0.19		0.27		0.47	
l		-68.13		-63.87		-50.63	
p		4		6		7	
AIC		144.27		139.74		115.26	

5.2.4 *Models 4* and *5*: Ferry competition

Attempting to better understand the influence of ferry competition, two different approaches were considered:

- the separation of low and high distances by determining a cut-off distance value from which the coefficient γ is allowed to assume a different value, and

- the addition of an explanatory variable corresponding to the factor between ferry and air travel time.

As mentioned in Section 4, there are several ferry companies operating in the Greek islands under study. The offered services change seasonally and were also affected by the Covid-19 pandemic, which makes the analysis much more complex. The approach of setting a cut-off distance f below which the ferry competition is more significant, seems suitable to this particular case study and is much simpler than embarking on a rigorous study of the ferry services.

Bearing this in mind, the distance variable was separated in Dl_{ij} and Dh_{ij} regarding lower and higher distances, respectively, resulting

$$V_{ij} = k\, Dl_{ij}^{\gamma_l}\, Dh_{ij}^{\gamma_h}\, P_i^{\delta}\, P_j^{\epsilon}\, T_i^{\theta}\, T_j^{\iota}\, S^{\lambda}\ , \tag{5.4}$$

$$\text{where} \quad Dl_{ij} = \begin{cases} d_{ij} & \text{if } d_{ij} \leq f \\ 1 & \text{otherwise} \end{cases} \ , \quad Dh_{ij} = \begin{cases} d_{ij} & \text{if } d_{ij} > f \\ 1 & \text{otherwise} \end{cases} \ .$$

Within the Dodecanese archipelago, there is a fair amount of companies offering routes assuring connectivity with Rhodos. Samos island is also served, as it has two ports, that serve many different companies, focused on different regions, working as a hub. However, for longer travels, like MJT-RHO ($d_{ij} = 323km$) and LXS-RHO ($d_{ij} = 465km$), passengers have to use a combination of ferry companies, like *Blue Star Ferries* to arrive to Samos and *Dodekanisos Seaways* to travel from Samos to Rhodes, which is extremely inconvenient. For this reason, the cut-off distance f was set to 323 km.

The results of the linear regression of the logarithmic transformation of (5.4), with $f = 323km$ are presented in Table 5.7. As expected, $\gamma_l > \gamma_h$, as ferry competitiveness is more significant for lower distances. Although the R^2 value is much higher for *Model 4* than for *Model 3*, the distance cut-off caused some unwanted effects:not only do the population variables present less significant estimates, but also their coefficients values change and adopt negative values, which does not go in line with the reasoning of that explanatory variable. This phenomenon is explained by the high correlation between distance and population in this particular case study.

In order to validate this assumption for the value of f, this value was iterated, so that the resultant values of R^2 could be analyzed, as shown in Figure 5.2. It is interesting to note that the graph presents exactly 9 stages, corresponding to the distances d_{ij} of the 9 links considered to the calibration. As expected, the best value of R^2 corresponds to distinguishing links from the Dodecanese archipelago plus Samos, from the other North Aegean, more distant, islands with worse sea connectivity to Rhodes. However, the other high plateau was not expected, as KGS and KZS present similar distances and sea travel times to RHO.

Table 5.5 presents the 9 links considered to the calibration ordered from nearest to farthest from RHO. Sea travel times, though not linearly, increase with distance, whereas air travel times present much

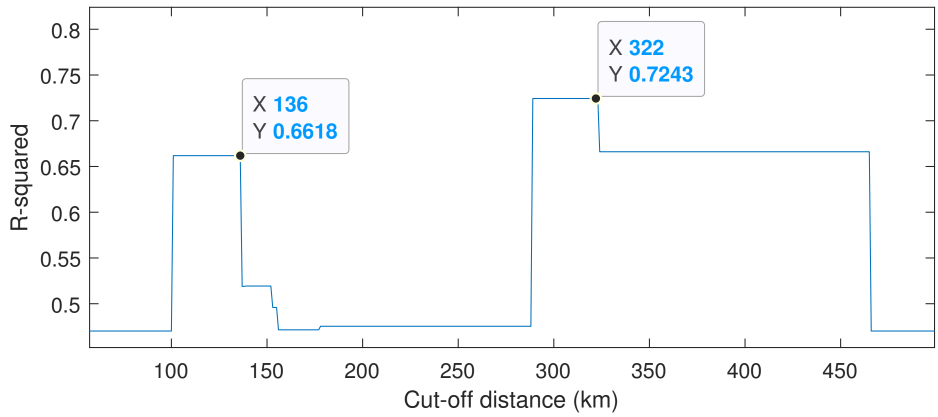


Figure 5.2: R^2 value in function of cut-off distance value f.

more irregular variations, which affect the competitiveness between these two transportation modes. Therefore, one can infer that distance, though correlated, does not represent correctly ferry competition. Thus, a new variable to serve as a proxy for ferry competition should be considered.

Table 5.5: Distance and both sea and air travel time to/from RHO.

Airport	d (km)	tt_{sea} (hours)	tt_{air} (hours)	F_{ij}
KGS	100	2.50	4.00	0.625
KZS	136	3.50	0.500	7.00
AOK	138	4.50	0.670	6.75
KSJ	152	4.50	1.50	3.00
JTY	155	9.50	3.25	2.92
SMI	177	8.50	0.750	11.3
HER	288	12.5	1.00	12.5
MJT	465	14.5	2.75	5.27
LXS	323	22.5	3.50	6.43

Although it is a simplistic approach, ferry competitiveness was represented by a factor between ferry and air travel time,

$$F_{ij} = \frac{tt_{sea}}{tt_{air}} \, ,$$

with tt_{sea} and tt_{air} standing for travel time by sea and air, respectively, i.e. the lower the value, the higher the ferry competitiveness when comparing to air transport. Adding this explanatory variable results in the following model

$$V_{ij} = k \, d_{ij}^{\gamma} \, P_i^{\delta} \, P_j^{\epsilon} \, T_i^{\theta} \, T_j^{\iota} \, S^{\lambda} \, F_{ij}^{\mu} \; . \tag{5.5}$$

The results of the linear regression of the logarithmic transformation of (5.5), presented in Table 5.7, seem to indicate a good model, however one should be wary about the high correlation between F_{ij} and other variables, namely d_{ij}, to avoid being misguided in the presence of multicollinearity. Table 5.6 shows the correlation coefficients between F_{ij} and the other explanatory variables and their respective p-values, i.e. the probability of the null hypothesis that there is no relationship between the two variables.

Table 5.6: Correlation coefficients of $\log(F_{ij})$ with the other explanatory variables and respective p-values.

Variable	Correlation Coefficient	p-value
$\log(d_{ij})$	0.5480	$<$ 0.0001
$\log(P_i)$	0.1517	0.1735
$\log(P_j)$	0.1583	0.1555
$\log(T_i)$	-0.2053	0.0643
$\log(T_j)$	-0.2084	0.0602
$\log(S)$	-0.0491	0.6613

As expected, there is a positive correlation between F_{ij} and d_{ij} since the air travel time increases linearly with d_{ij}, while sea travel time increases steeply for island pairs further apart, because of the multiple stops and/or vessel changes. This significant correlation leads to a relevant change in distance coefficient value γ and an unwanted increase of its p-value, as shown in Table 5.7. The correlation between F_{ij} and the touristic attractiveness of origin and destination islands, T_i and T_j, is not so obvious, but it makes sense for it to take a negative coefficient, since more touristic pairs of islands have better ferry services, therefore lowering the sea travel time and consequently F_{ij}.

5.2.5 ***Model 6*: Covid-19 pandemic**

The Covid-19 pandemic represents an important exogenous shock, similar to the ones Chevallier et al. [9] accounted for. Therefore, a dummy variable C was included, with $C = 10$ for August and November 2020, and $C = 1$ for 2018 and 2019 (following the same explanation as for S, in Section 5.2.3), yielding

$$V_{ij} = k\, d_{ij}^{\gamma}\, P_i^{\delta}\, P_j^{\epsilon}\, T_i^{\theta}\, T_j^{\iota}\, S^{\lambda}\, F_{ij}^{\mu}\, C^{\nu} \ . \tag{5.6}$$

The results of the linear regression of the logarithmic transformation of (5.6) are presented in Table 5.7. In spite of the slightly higher R^2, *Model 6* AIC increased when compared to *Model 5*, indicating that this extra variable C increases the complexity of the model without conferring a satisfactory improvement. Although the Covid-19 pandemic clearly influenced air travel passenger volume, as shown in Figure 5.1, the coefficient μ presented a high p-value of 0.38, i.e. a low level of significance. This 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.

5.2.6 ***Model 7*: Moderation according to Covid-19 pandemic**

As a preliminary study, data from 2018 and 2019 was used to obtain *Pre-Model (a)*, regarding passenger demand before the pandemic, and data from 2020 to obtain *Pre-Model (b)* for the period during the pandemic. This approach was not considered suitable to predict passenger demand because there is no sufficient data, only 55 data points for *(a)* and 27 for *(b)*. However, it led to the *Model 7* idea, and therefore is worth being referred to.

Table 5.7: *Models 4*, *5*, and *6* coefficient estimates and p-values.

		Model 4 (5.4), with $f = 323$		*Model 5* (5.5)		*Model 6* (5.6)	
Coef.	Var.	Estimate	p-value	Estimate	p-value	Estimate	p-value
$\log(k)$	-	-12	<0.01	-5.2	<0.01	-5.2	<0.01
γ	d_{ij}	-	-	0.67	0.13	0.69	0.12
γ_l	Dl_{ij}	4.3	<0.01	-	-	-	-
γ_h	Dh_{ij}	3.7	<0.01	-	-	-	-
δ	P_i	-0.023	0.69	0.17	<0.01	0.17	0.01
ϵ	P_j	-0.013	0.82	0.19	<0.01	0.19	<0.01
θ	T_i	0.93	<0.01	0.86	<0.01	0.86	<0.01
ι	T_j	0.90	<0.01	0.75	<0.01	0.74	<0.01
λ	S	0.44	<0.01	0.38	<0.01	0.37	<0.01
μ	F_{ij}	-	-	0.73	<0.01	0.73	<0.01
ν	C	-	-	-	-	-0.086	0.38
R^2		0.72		0.60		0.61	
l		-23.84		-38.88		-38.44	
p		8		8		9	
AIC		63.68		93.75		94.88	

The moderation of the explanatory variables according to the Covid-19 pandemic is a good compromise between *Model 6* and designing two separate models, as the effect of most explanatory variables will be estimated taking all 82 data points into account. The moderation of all explanatory variables would result in

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

with $C = 0$, for the period before the pandemic and $C = 1$ for the period during the pandemic. Table 5.8 presents the coefficients and respective p-values that would be calibrated for such model. It can be observed that distance coefficient γ presented a higher value in *Model 5* (0.69), while still integrating the periods before the pandemic and during the pandemic, than when separating these two periods, it took a lower value for the period before the pandemic ($\gamma = 0.37$) having an heavier effect during the pandemic ($\gamma + \gamma' = 0.37 + 0.90$). Several variables present a high *p-value*, thus these variables were disregarded one by one by decreasing p-value value order, according to a backward stepwise procedure, until achieving an appropriate model, *Model 7*:

$$V_{ij} = k\, d_{ij}^{C\gamma'}\, P_i^{\delta}\, P_j^{\epsilon}\, T_i^{(\theta + C\theta')}\, T_j^{(\iota + C\iota')}\, S^{\lambda}\, F_{ij}^{(\mu + C\mu')}\,. \tag{5.7}$$

The results of the linear regression of the logarithmic transformation of (5.7) are presented in Table 5.8. Tourism, T_i and T_j, continues to have a positive effect on passenger demand, both before the pandemic ($\theta > 0$ and $\iota > 0$) and during the pandemic ($\theta + \theta' > 0$ and $\iota + \iota' > 0$), but less prominent during the pandemic period ($\theta + \theta' < \theta$ and $\iota + \iota' < \iota$). 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.

Table 5.8: *Model 7* coefficient estimates and p-values.

		Moderation of all exp. var.		*Model 7* (5.7)	
Coef.	Var.	Estimate	p-value	Estimate	p-value
$\log(k)$	-	-5.3	<0.01	-4.2	<0.01
γ	d_{ij}	0.37	0.42	-	-
δ	P_i	0.19	0.02	0.20	<0.01
ϵ	P_j	0.23	<0.01	0.22	<0.01
θ	T_i	0.95	<0.01	0.86	<0.01
ι	T_j	0.81	<0.01	0.75	<0.01
λ	S	0.36	<0.01	0.35	<0.01
μ	F_{ij}	0.94	<0.01	1.0	<0.01
γ'	d_{ij}	0.90	0.05	0.76	0.010
δ'	P_i	-0.047	0.74	-	-
ϵ'	P_j	-0.011	0.42	-	-
θ'	T_i	-0.28	0.22	-0.31	0.11
ι'	T_j	-0.15	0.50	-0.27	0.17
λ'	S	-0.013	0.95	-	-
μ'	F_{ij}	-0.66	0.04	-0.72	0.02
R^2		0.65		0.64	
l		-34.25		-35.25	
p		15		11	
AIC		98.51		92.50	

5.2.7 Model Specification

From the first set of three models presented in Table 5.4, *Model 3* clearly stands out, as it confers the better fit, with the higher R^2 of 0.47, without being unjustifiably more complex, as it shows the lowest AIC value of the three. Table 5.9 presents the correlation coefficients between the explanatory variables. The correlation coefficients are, as expected, very similar for combinations with the pairs P_i P_j and T_i T_j, therefore there is only one entry for each pair. Note that the last column and last line do not refer to *Model 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. This model is a good start as it is reasonably adapted to this case study but still suitable for other touristic regions with different characteristics.

The models from the second set, presented in Table 5.7, are more tailored to this specific case study as they consider the ferry boat competition. However, when the Covid-19 pandemic is added as an extra factor, in *Model 6*, this variable shows to be not significant and the AIC increases. *Model 7*, in Table 5.8, allows the effect of some explanatory variables to be moderated by the pandemic, further lowering the

Table 5.9: Correlation coefficients and respective p-values (in brackets).

	$\log(d_{ij})$	$\log(P)$	$\log(T)$	$\log(S)$	$\log(F)$
$\log(d_{ij})$	-	0.28 (0.01)	-0.42 (<0.01)	-0.12 (0.29)	0.55 (<0.01)
$\log(P)$	0.28 (0.01)	-	0.26 (0.02)	0.04 (0.71)	0.16 (0.16)
$\log(T)$	-0.42 (<0.01)	0.26 (0.02)	-	0.07 (0.54)	-0.2 (0.06)
$\log(S)$	-0.12 (0.29)	0.04 (0.71)	0.07 (0.54)	-	-0.05 (0.66)
$\log(F)$	0.55 (<0.01)	0.16 (0.16)	-0.2 (0.06)	-0.05 (0.66)	-

AIC. Therefore, *Model 7*, with $R^2 =$ 64%, is the model that better describes the passenger demand for this case study.

When interpreting *Model 7*, it is important to acknowledge the correlation between variables d_{ij} and F (presented in Table 5.9) when analyzing the effect of these variables on passenger demand V_{ij} within the pandemic context. A positive correlation between these two values was expected, as described in Section 5.2.4. 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*.

5.3 Scenarios Estimation

Different sets of scenarios were obtained using models presented in Section 5.2. This section covers the translation of the predictions from a gravity model to a set of scenarios to give as input to the Route and Flight Schedule Optimization model. Note that the main contribution of this work is to develop a robust Route and Flight Schedule Optimization model by accounting for the uncertainty inherent to passenger behavior and exogenous factors through the consideration of multiple passenger demand scenarios.

First, it is important to remark that the data used to obtain the models is limited, hence some assumptions were necessary. Figure 5.3 contributes to the readability of this paragraph. As explained in Section 5.1, the number of passengers flying on route f, that departs from i and arrives to j, V_{ij}, is used as a proxy for unconstrained demand from i to j, U_{ij}, i.e. it is assumed that $U_{ij} \approx V_{ij}$. However, this historical data is influenced by the offered service (offered routes, price, frequency, etc.), therefore it is not a perfect indicator of the passengers that desire to fly from i to j. The second assumption that importantly affects the estimations obtained is the fact that part of the passengers flying on route f, could be actually just using it as a leg of a longer itinerary, i.e. their origin could be different from i and/or their destination different from j (passengers traveling from i to k in Figure 5.3 example, U_{ik}). Thereby, the number of passengers flying route f, V_{ij} could be considerably inflating what is interpreted as the number of passengers desiring to travel from i to j, U_{ij}. This second assumption leads to an important overestimation of passenger demand.

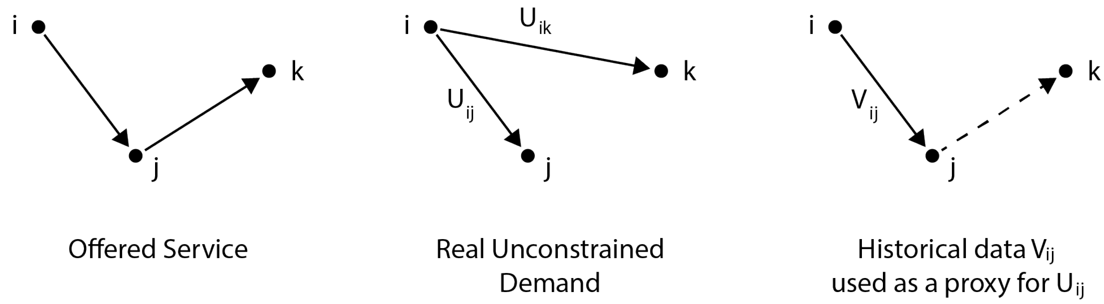


Figure 5.3: Use of observed passenger volume as a proxy for passenger demand.

Note that, although the estimations resultant from using this calibration data are affected, the applicability of the models themselves is not, as they would perform better with more suitable data. To overcome the overestimation of the passenger demand volume, the results were normalized using an artificial constant. This scaling factor was computed so that: i) the number of observed passengers (all traveling to and from RHO), denoted by v_{obs} and ii) the number of estimated passenger demand between airports from different PSO groups, denoted by v_{inter}, are in the same order of magnitude. This rationale rises from the fact that passengers traveling within the same PSO group, i.e. links between AOK and KSJ or among JKL, JTY, KGS, and LRS, certainly would not need to stop in RHO and therefore were not accounted for in the observed data. Thus, the estimated passenger demand scaling factor is given by:

$$SF = \frac{v_{inter}}{v_{obs}} \ , \tag{5.8}$$

where the numerator v_{inter} corresponds to the inter-group estimated passenger demand according to the median of the MC iterations for the sum of every route, and both numerator v_{inter} and denominator v_{obs} refer to an equal time period. This rationale ensures that the number of passenger demand estimated for the whole network is not under-estimated either. The values assumed by SF when using *Model 3* and *Model 7* will only be presented in the next section, after further describing how the estimates were obtained through MC simulations.

5.3.1 Monte Carlo Simulation

Although the procedures to obtain the passenger demand scenarios estimates using *Model 3* or *Model 7* are very similar, following the rationale explained in Section 3.1.3, the latter is a little more complex, as it considers the moderation of the pandemic context effect. Therefore, the procedure performed while using *Model 3* will be explained first, and just then the additional methodological steps performed while using *Model 7* will be explained.

Model 3

To obtain the *Model 3* passenger demand volume estimations for every link between the 8 airports under study for both high and low seasons, batches of $b = 100$ iterations were defined and $k = 4,900$ loops were performed, resulting in a total of $N = 490,000$ iterations, according to the method described in Section 3.1.3. These estimations are presented in Figure 5.4, that shows the distribution of weekly passenger demand estimations for each one of the 56 links before normalization.

These estimates were normalized through the dividing by the scaling factor, SF, described previously. The total number of passengers observed arriving and departing RHO was $10,736$ for all the three years (2018, 2019, and 2020), both August and November. The estimated median of monthly total sum of inter-group passengers' demand for high plus low season was $9,220$, as the total sum consists on $9,353$ (as observed in Figure 5.5, $6,542 + 2,811$) where 133 of this corresponds to trips within the same PSO group. Thus, according to (5.8), $SF = \frac{9220}{10736/3} = 2.5764$. Figure 5.6 is the equivalent of Figure 5.5 after normalization.

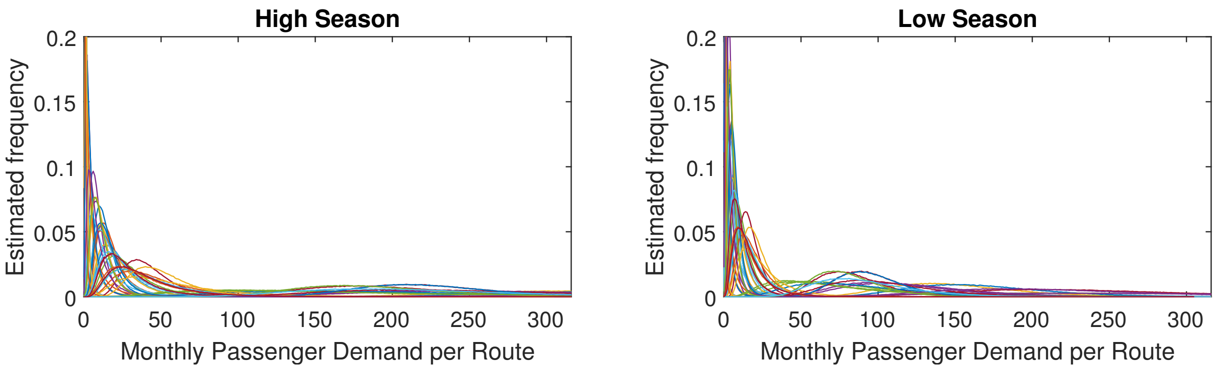


Figure 5.4: Monthly passenger demand per link estimated using *Model 3* iteration frequency before normalization.

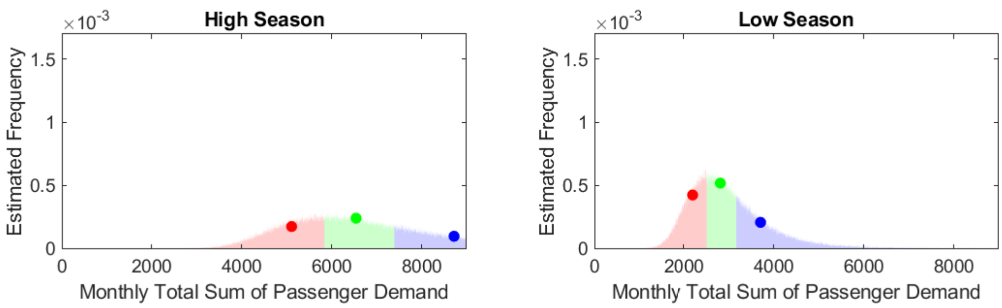


Figure 5.5: Monthly total passenger demand estimated using *Model 3* iteration frequency before normalization.

Model 7

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

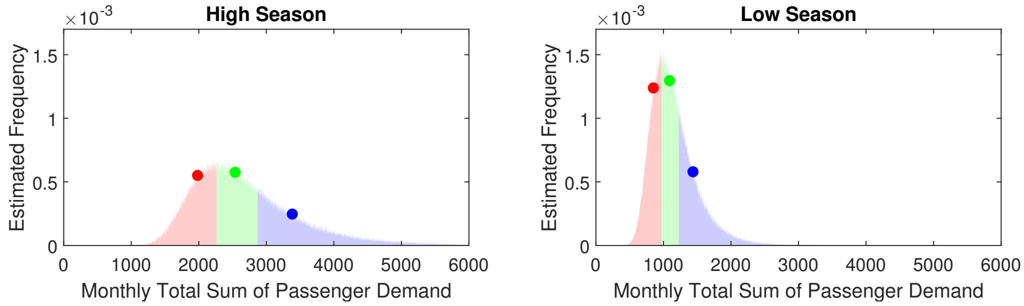
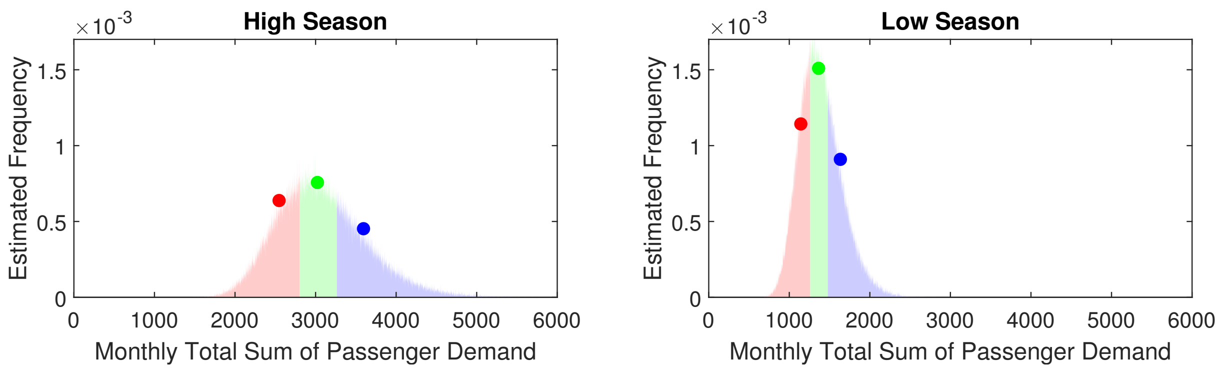


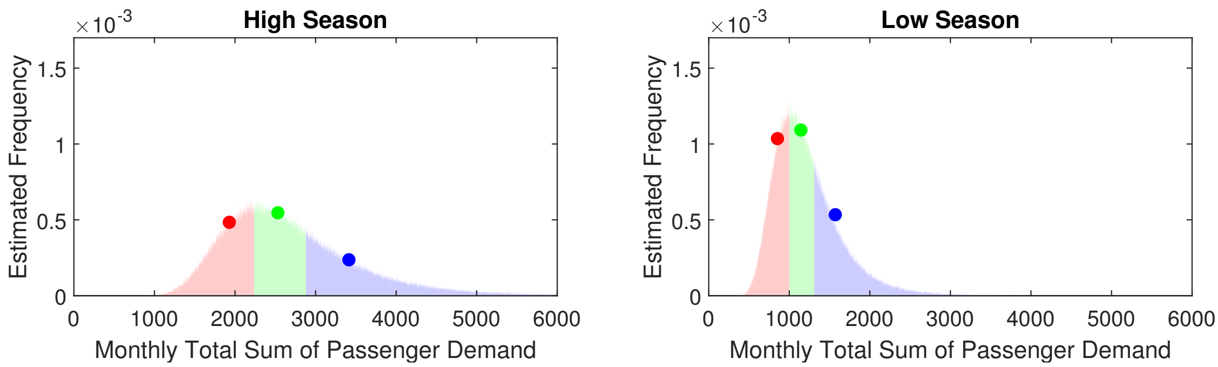
Figure 5.6: Monthly total passenger demand estimated using *Model 3* iteration frequency after normalization.

Figure 3.2, Section 3.1.3, so that the probability of occurring pandemic context, p_C (probability of $C = 1$), was considered. In each iteration, a random number from a binomial distribution following the probability p_C is generated, and V_{ij} is calculated according to the resultant value of $C = 0$ or $C = 1$.

A probability of $p_C = 1/3$ was used to obtain a value for the numerator used to compute the scaling factor (5.8), since one of the three years of observations referred to a period during the pandemic, resulting in the value $\frac{4950}{10736/3} = 1.3832$. This simulation consisted on $N = 790,000$ iterations, under the same MC stopping rule described in Section 3.1.3. Figure 5.7 presents the normalized distribution of the total sum of passengers demand estimations for periods outside ($p_C = 0$) and within ($p_C = 1$) a pandemic context (MC simulations with $N = 210,000$ and $N = 620,000$ iterations, respectively).



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



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

Figure 5.7: Monthly total passenger demand estimated using *Model 7* iteration frequency, outside and within a pandemic context.

5.3.2 Demand Scenarios Selection

As explained in Section 3.1.3, the distribution of the total sum of passengers, shown in Figure 5.6 for *Model 3* and Figure 5.7 for *Model 7*, was used to define three scenarios: i) the most likely, ii) the pessimistic, and iii) the optimistic scenarios; in order to take the covariance between the different links passenger demand into account.

The most likely scenario assumes one of the iterations corresponding to the median of the total sum of passengers (highlighted in these figures with green dots), selected according to the criterion described in Section 3.1.3. Table 5.10 shows the most likely scenario when using *Model 3*, for high and low seasons, as an example. 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 (highlighted with red and blue dots in Figure 5.6, 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). Table 5.11 shows the pessimistic and optimistic scenarios, for low season, as an example. Note that, for each route, the passenger demand estimated is not necessarily lower for the pessimistic scenario compared to the optimistic one, which only happens because the covariance between the different links passenger demand is taken into account.

Table 5.10: Monthly passenger demand estimated for the most likely scenario for high/low seasons, using *Model 3*.

Dep. \ Arr.	AOK	JKL	JTY	KGS	KSJ	KZS	LRS	RHO
AOK	-	10 / 5	8 / 3	137 / 59	2 / 0	67 / 31	10 / 5	92 / 41
JKL	7 / 4	-	0 / 0	2 / 0	5 / 3	11 / 7	0 / 0	11 / 6
JTY	7 / 3	0 / 0	-	8 / 3	4 / 2	15 / 9	0 / 0	18 / 9
KGS	152 / 63	3 / 1	10 / 3	-	111 / 45	248 / 109	6 / 2	223 / 84
KSJ	2 / 0	8 / 4	6 / 2	107 / 44	-	58 / 26	7 / 4	81 / 35
KZS	80 / 35	19 / 9	22 / 10	268 / 114	64 / 38	-	17 / 9	102 / 39
LRS	7 / 4	0 / 0	0 / 0	3 / 2	5 / 3	10 / 6	-	11 / 7
RHO	89 / 40	16 / 8	21 / 10	194 / 78	73 / 33	83 / 34	16 / 8	-

Note that the pessimistic and optimistic scenarios do not correspond necessarily to fewer/more passengers in each individual route. In fact, one can verify in Table 5.11 that, for many lower demand routes, the number of passengers estimated for the pessimistic scenario is higher than the for the optimistic one, which does not happen for the higher demand routes.

Table 5.11: Monthly passenger demand estimated for the most pessimistic/optimistic scenarios for low season, using *Model 3*.

Dep. \ Arr.	AOK	JKL	JTY	KGS	KSJ	KZS	LRS	RHO
AOK	-	6 / 3	4 / 3	43 / 78	1 / 0	24 / 44	6 / 3	32 / 43
JKL	5 / 2	-	0 / 0	1 / 0	3 / 2	7 / 5	0 / 0	7 / 3
JTY	3 / 2	0 / 0	-	3 / 3	2 / 2	7 / 9	0 / 0	9 / 7
KGS	48 / 83	1 / 0	5 / 3	-	35 / 64	72 / 187	3 / 1	63 / 108
KSJ	1 / 0	4 / 3	3 / 2	32 / 63	-	20 / 40	4 / 2	27 / 41
KZS	26 / 50	9 / 8	10 / 11	71 / 200	21 / 43	-	9 / 7	30 / 54
LRS	4 / 2	0 / 0	0 / 0	2 / 1	3 / 2	6 / 5	-	7 / 4
RHO	33 / 42	9 / 5	10 / 8	57 / 98	27 / 37	28 / 46	9 / 5	-

Chapter 6

Implementation of Route and Flight Schedule Optimization

In this chapter the problem formulation presented in Section 3.2 is tested with some illustrative examples, in Section 6.1, and then applied to the Rhodes based PSO network case study. Section 6.2.1 presents the input constants defined according to the case study. Section 6.2.2 elaborates on two ways of relaxing the problem for the sake of tractability.

6.1 Illustrative Examples

A few illustrative examples with 4 ($NA = 4$) to 6 airports are hereafter presented, in order to validate the developed optimization program and assess the tractability of the problem. The problem is solved on an Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz with 16 GB of RAM using the optimizer version 35.01.01 of IBM's FICO Xpress software. The passenger demand used as input is presented in Table 6.1, where airport 4 is defined as the hub airport. Considering time steps of half an hour, the travel time, t_f was set to be two time steps between every airport except between 2 and 3, which are only one time step travel time apart. Two airplanes were considered: one of type 1 with 48 seats ($z(1) = 1$, $s(1) = 48$), the other of type 2 with 37 seats ($z(2) = 1$, $s(2) = 37$).

Table 6.1: V_{ij} values for three different scenarios used as an illustrative example.

Dep. \ Arr.	1	2	3	4 (hub)
1	-	27 / 34 / 46	3 / 3 / 1	20 / 23 / 23
2	27 / 38 / 48	-	0 / 0 / 0	43 / 56 / 58
3	2 / 2 / 2	0 / 0 / 0	-	4 / 3 / 2
4 (hub)	19 / 22 / 26	40 / 49 / 62	4 / 4 / 2	-

In order to validate the multiple day approach, for the 1st illustrative example, the program was run for 2 days ($ND = 2$) with 4 operational hours each ($NT = 8$), corresponding to 8 a.m. until 12 a.m..

An optimal solution was computed within 3.9 seconds. The depth of the solutions found is shown in Figure 6.2 by the blue triangles and is also visible in the BB-Tree in Figure 6.1, where solutions are represented by the green squares. The improvement of the optimality gap can be analyzed in Figure 6.2, as the solutions found assume better absolute objective values and the best bound increases through the search process.

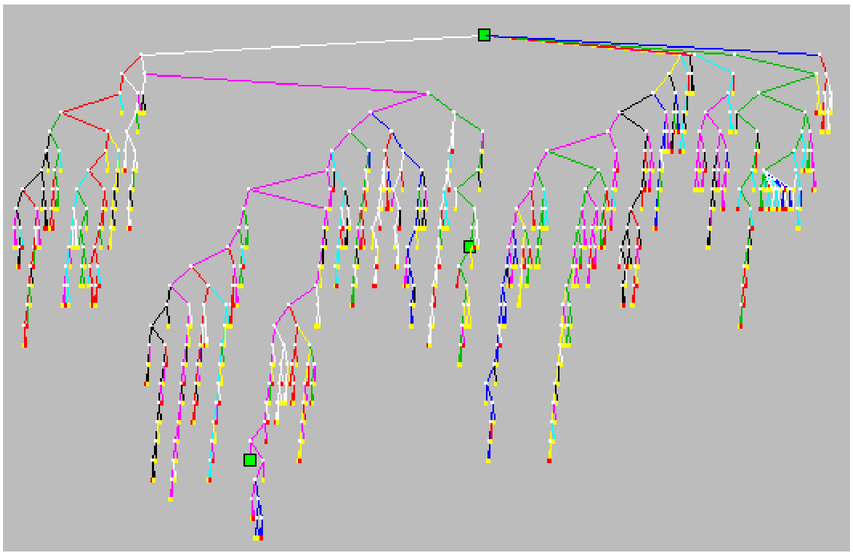


Figure 6.1: 1st Illustrative example BB-Tree diagram.

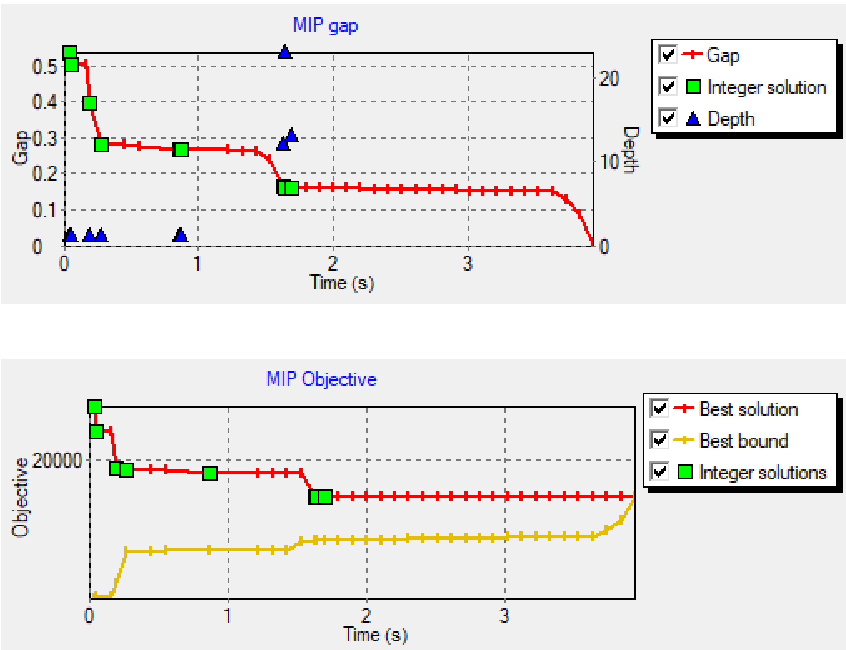
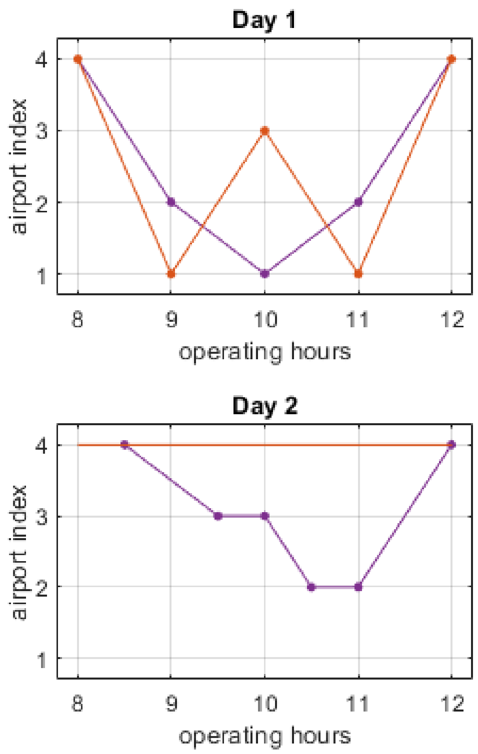
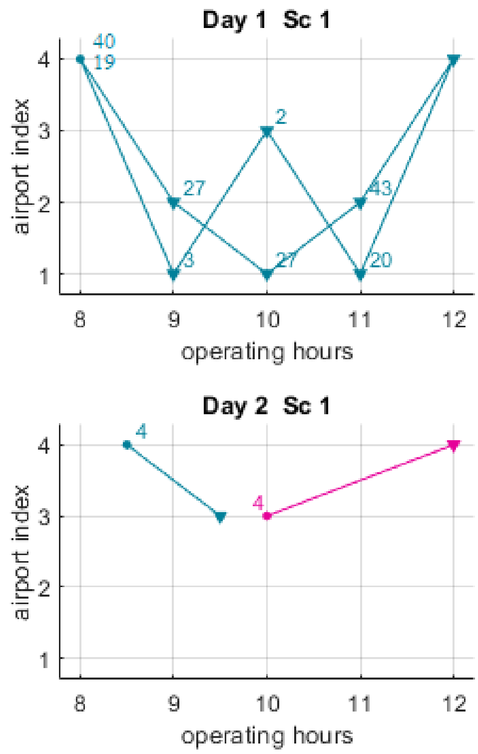


Figure 6.2: MIP search optimality gap, solutions depth, and objective evolution (1st illustrative example).

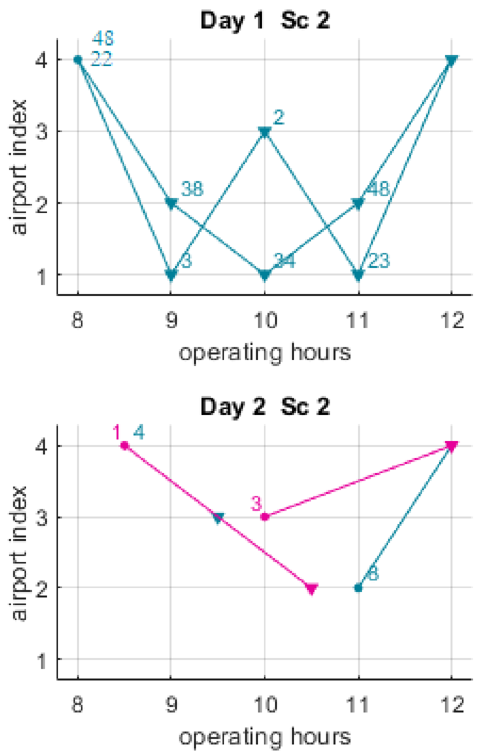
Figure 6.3 shows the resulting network schedule and carried passengers, where it is possible to confirm that every constraint described in Section 3.2.1 is met. Figure 6.3(a) represents the scheduled flights for aircraft type 1, in purple, and type 2, in orange (note time-step continuity and at-hub maintenance). Figures 6.3(b), (c), and (d), show the number of passengers traveling in direct routes, in blue, and one stop itineraries, in pink (note that passenger demand is satisfied for each scenario).



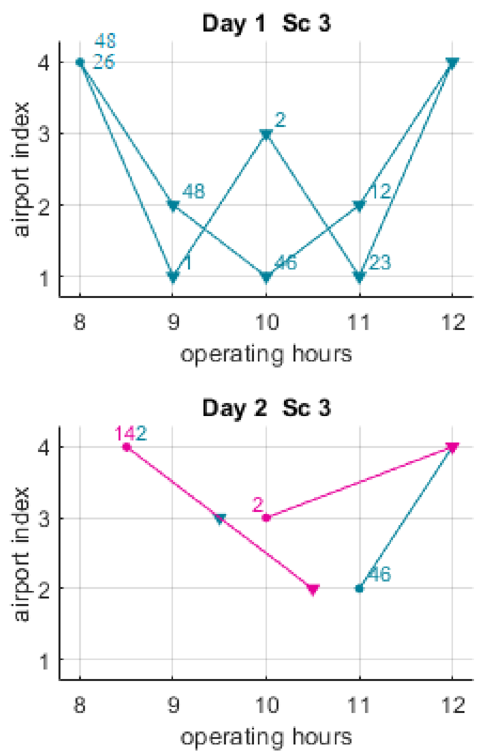
(a) Scheduled flights.



(b) Passengers carried in scenario 1.



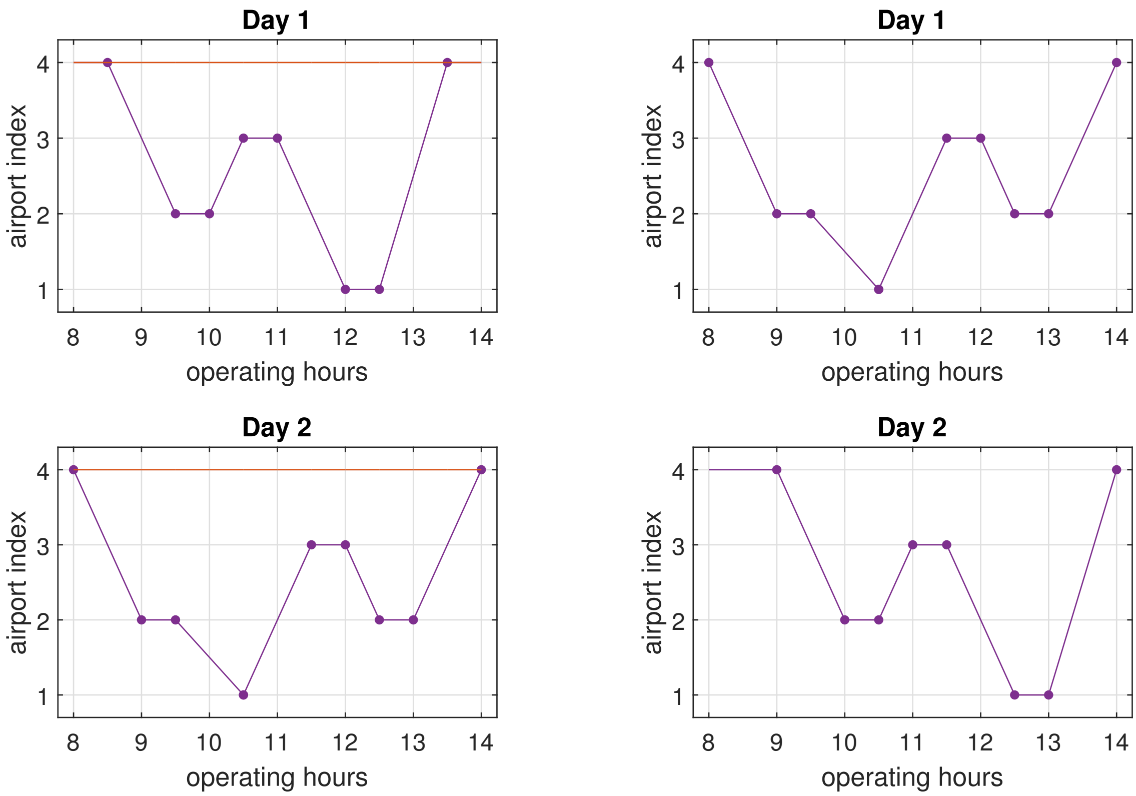
(c) Passengers carried in scenario 2.



(d) Passengers carried in scenario 3.

Figure 6.3: Optimized network schedule and passengers carried in each flight for different scenarios.

When keeping all previous options (passenger demand to be satisfied in each scenario, aircraft available, etc.), except for increasing the operational time period to be considered, the complexity of the problem increases significantly. In the 2nd illustrative example, just by increasing the number of operational hours from 4 to 6 per day ($NT = 12$), the program took 357.4 s to achieve the optimal solution, presented in Figure 6.4 (a). In the 3rd example, running the same problem but with only one aircraft of type 1 ($z(1) = 1$, $s(1) = 48$), the computation time was reduced to 20.4 s, resulting in the solution in Figure 6.4 (b). Note that solutions (a) and (b) present exactly the same two flight patterns as the aircraft leaves and returns to the hub, just arbitrarily shifted in time, switching days. This happens because there are several equally optimal solutions, one of which is output by the optimization program.



(a) two aircraft

(b) one aircraft

Figure 6.4: Route and Flight Schedule Optimization solution for 2nd and 3rd illustrative examples, respectively.

In the 4th example, the same options from the 1st illustrative example were kept ($ND = 2$, $NT = 8$, with the same two aircraft), except for adding one island ($NA = 5$), and the program took 7.7 s to achieve a zero optimality gap (the values considered for passenger demand for the additional links are of the same order of magnitude). Increasing the number of airports to $NA = 6$, maintaining $ND = 2$ and $NT = 8$, in the 5th example, leads to an unfeasible problem, so those would have to be adjusted. Adjusting $NT = 12$, 6th example, the complexity of the problem drastically increased, as after 24h, the optimality gap was still 21% (Figure 6.5).

By analyzing the previous examples, summarized in Table 6.2, one can infer that increasing NA by itself might not increase significantly the complexity of the problem (compare 1st and 4th examples), but

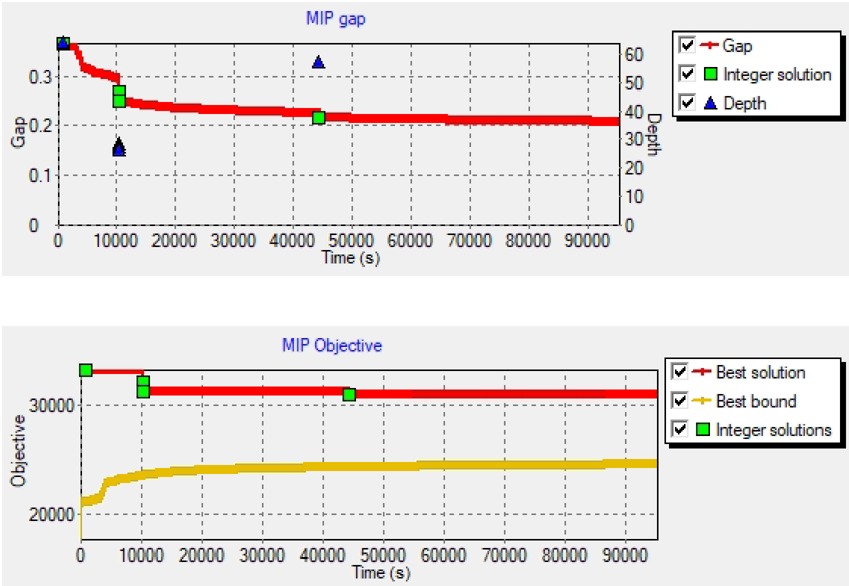


Figure 6.5: MIP search optimality gap, solutions depth, and objective evolution (6th illustrative example).

Table 6.2: Illustrative examples summary.

Illustrative example	NA	ND	NT	nº aircraft	number of links with $V_{ij} \neq 0$	computation time
1st	4	2	8	2	10	3.9 s
2nd	4	2	12	2	10	357.4 s
3rd	4	2	12	1	10	20.4 s
4th	5	2	8	2	16	7.7 s
5th	6	2	8	2	24	no feasible solution
6th	6	2	12	2	24	21% optimality gap after 24 h

increasing the number of time steps to consider NT and/or ND (1st *vs.* 2nd), or the number of aircraft available does (2nd *vs.* 3rd). However, when the number of airports to be connected increases, the time period to consider and/or the number of aircraft to be available have to be adjusted (5th example) and consequently the complexity of the problem can increase drastically (4th *vs.* 6th). Therefore, it is expected that the application of the optimization program to the case study will be computationally very demanding, hence the need for problem relaxation.

6.2 Case Study Application

6.2.1 Input Constants Estimation

The case study comprises of connections among 8 islands ($NA = 8$). Different sets of possible direct routes, one, two and three stops itineraries were pre-computed and fed as input to the program according to the cluster relaxation explained later in Section 6.2.2 (i.e., constants NF, NG, NH, NI, and parameters $a_f(f)$, $a_g(g)$, $a_h(h)$, $a_i(i)$).

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 ($NT = 32$). It is beneficial to have the least number of operational days per week, ND, possible, as it allows a reduction of operational costs and also reduces the complexity of the optimization problem. Solutions in Chapter 7 use $ND = 1$ or $ND = 2$.

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

A higher number of scenarios would also increase considerably the complexity and, consequently, the tractability of the problem. Therefore, the $NS = 3$ scenarios seem sufficient to test the robustness that this approach can add to the solutions, as the most important point to prove is that the resulting network is capable of handling a range of passenger demand scenarios, from pessimistic to optimistic extremes.

The case study network will be optimized assuming the aircraft currently used are maintained. As such, there are $NR = 2$ types of aircraft, presented in Table 6.3 (as half an hour time steps are used, $c_F(r)$ stands for the cost of flying the airplane type r for half an hour).

Table 6.3: Aircraft related parameters.

aircraft type r	number of available aircraft $z(r)$	seat capacity $s(r)$	cost of flying $2 \times c_F(r)$
ATR 42-320 ($r = 1$)	2	48	1502 €/h
Dash 8-100 ($r = 2$)	1	37	1502 €/h

The cost of flying consists on the cost for the airline of operating a flight with that aircraft type. Therefore, constant c_F was estimated including fuel, crew, airport and airspace fees, maintenance, etc. Estimates for all $c_F = 751$, $c_A = 75$ (as the parking cost is 150€/h for both aircraft types r and all airports a), $c_B = 5$, and $c_G = 5$ (as time cost for passengers was estimated to be 10€/h) were based on Leandro et al. [6], who used the Eurocontrol Standard inputs for cost-benefit analysis.

6.2.2 Problem Relaxation

As explained in 3.2.2, the problem, as presented before, is too complex and must be relaxed to ease tractability. Section 6.2.2 presents a method to reduce the number of possible direct routes, one, two, and three stops itineraries, whilst Section 6.2.3 presents a method to reduce the number of operational days to be considered simultaneously by the optimization program.

Clusters

By separating the 8 islands into clusters and defining restrictions for inter-cluster traveling, the number of possible itineraries is significantly reduced, maintaining the problem fairly close to the firstly introduced as it is a reasonable approximation.

The rationale to define said clusters consists on setting *a priori* each island as their own cluster and then joining a cluster with another whenever there is an airport from which there is less than half an hour of travel time to at least one airport from that second cluster. These pairs of airports with a time travel of less than half an hour and the resulting clusters are represented in Figure 6.6 by dashed lines and circles, respectively. Note that the resultant clusters coincide with the already existing PSO groups.

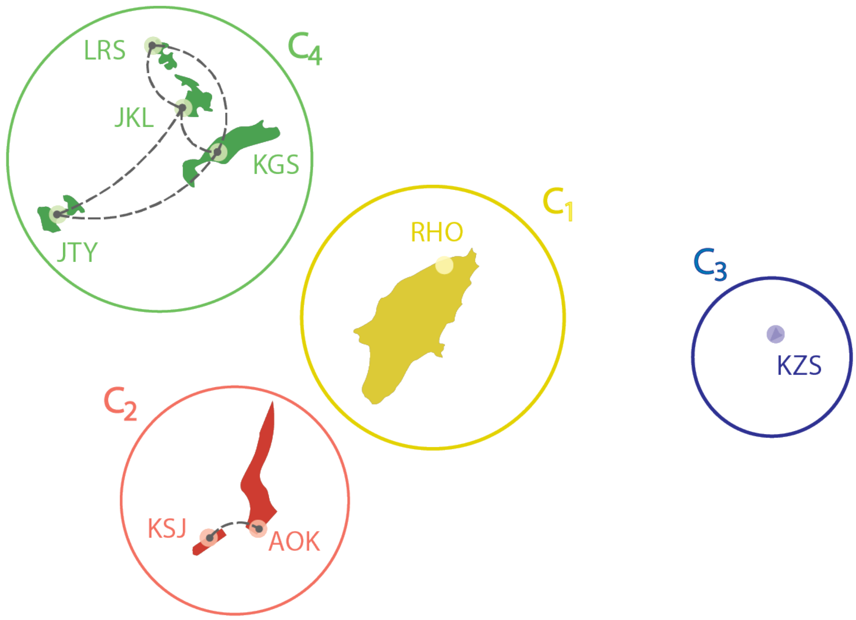


Figure 6.6: Cluster definition.

Once the clusters are defined, entry and exit airports can be chosen. For clusters C_1 and C_3 there is only one option, as the airport of their single island must be defined both as an entry and exit airport. For the two remaining clusters, C_2 and C_4, different combinations can be explored.

Firstly, one has to decide on the number of entry and exit airports for each cluster. Hereafter, in this section, the estimates obtained through *Model 3* will be used. As the number of passengers leaving a cluster is expected to be similar to the number of passengers entering it (as verified in Table 6.4), it only makes sense to have the same number of exit airports as of entry airports. If C_2 had two entry and two exit airports, it would be almost equivalent to two separate clusters, which goes against the intention of relaxing the problem, therefore there will be only one entry and one exit airports. On the other hand, C_4 includes 4 islands, allowing more options: one only entry/exit airport could be insufficient, as the passenger volume of all the four islands would have to pass by that airport; four entry/exit airports would be almost equivalent to defining four clusters, thereby is also inappropriate.

Afterwards, the entry and exit airports must be chosen. There could be two distinct entry and exit airports, as in Figure 6.7(a), where the routes flow in counterclockwise direction. However, passenger

Table 6.4: Between/within clusters monthly passenger volume, according to *Model 3* most likely scenario estimation for high + low season.

Dep. \ Arr.	C_1	C_2	C_3	C_4
C_1	0	235	117	351
C_2	249	4	182	419
C_3	141	207	0	468
C_4	369	425	415	43

volumes are estimated to be similar on both directions of the same link (as verified in Table 5.10), therefore the best choice for an entry airport is the same as for an exit airport. To choose the airports, heavier inter-cluster passenger volumes were prioritized, resulting in the remaining options in Figure 6.7 open to be explored. The sum of a high and a low season month volume of passengers estimated to be traveling inter-cluster with *Model 3* for each airport and each cluster (arriving plus departing) is presented in Table 6.5.

Table 6.5: Estimated monthly inter-cluster passenger volume per airport and cluster (*Model 3* most likely scenario estimation).

Airport	Inter-Cluster Passenger Volume	Cluster	Inter-Cluster Passenger Volume
RHO	1462	C_1	1462
AOK	959	C_2	1717
KSJ	758		
KZS	1530	C_3	1530
JKL	133	C_4	2447
JTY	149		
KGS	2036		
LRS	129		

The fewer the number of entry and exit airports, the larger the number of stops needed to fulfill the demand for inter-cluster links, only possible with non-direct itineraries (that are verified in every option except (d)). Note that supposing there can be connections between every cluster, configuration (b) would force that itineraries with up to at least two stops were considered in order to be possible to travel, for example, from JKL to KSJ.

Regarding the inter-cluster traveling restrictions, Figure 6.8 presents the different options considered. Configuration i) presents a network with RHO as the hub, where all inter-cluster traveling must pass through C_1. Since demand volume between C_3 and C_4 was higher, in ii) this link was also permitted. As of configuration (ii), traveling between C_2 and C_4 implies passing through C_1, which increases the number of stops in non-direct itineraries needed: for example, to travel from JKL to KSJ, itineraries with up to at least three stops have to be considered (in this case, stopping at KGS, RHO, and AOK). Therefore, configuration (iii) enables traveling between the two clusters with more islands, C_2 and C_4, in order to decrease the number of stops of non-direct itineraries. Finally, option (iv) allows traveling

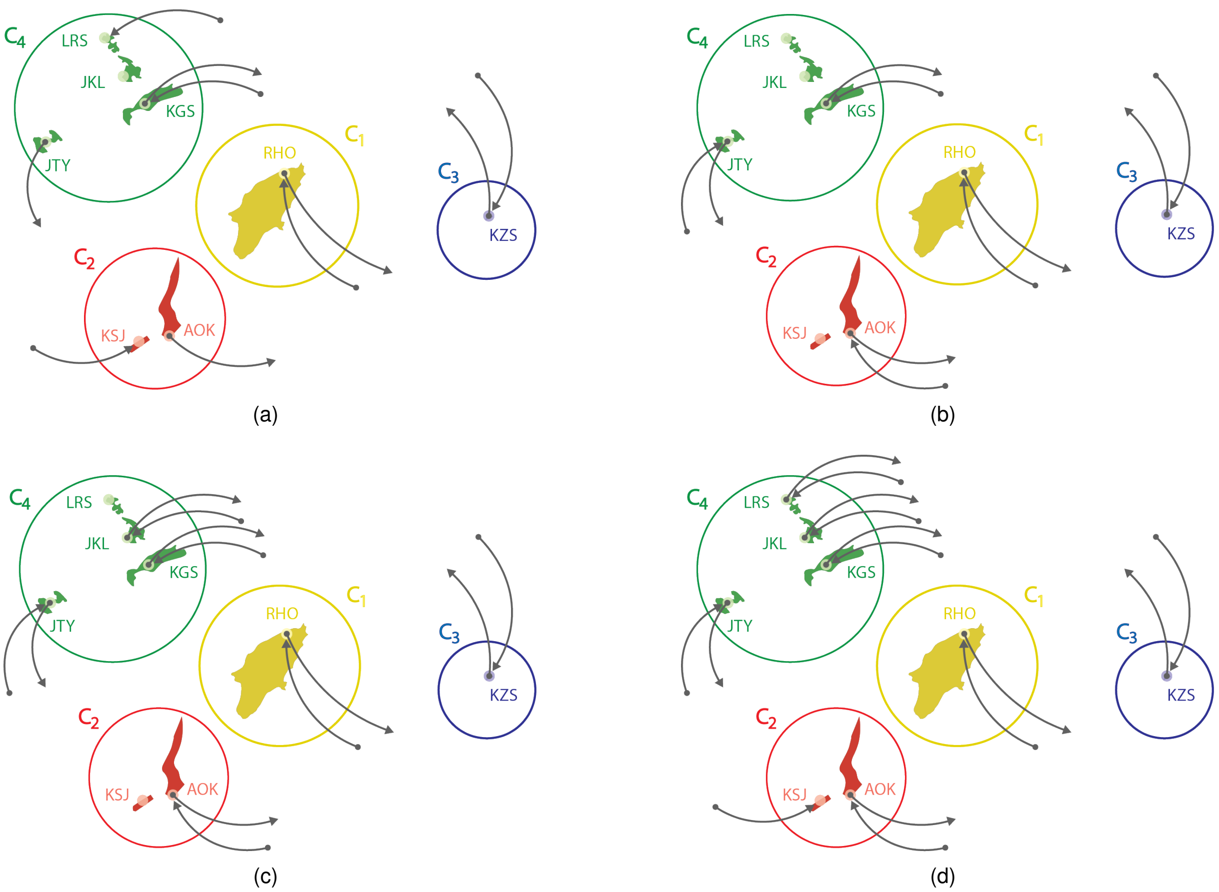


Figure 6.7: Different definitions for entry and exit airports within each cluster.

between any cluster.

Two additional restrictions were applied in order to relax the optimization problem: multiple stop itineraries exiting and returning to the same cluster were excluded and 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. Note that these two later restrictions by themselves when applied to a configuration d(iv) (that did not limit direct flights) already narrows significantly the number of possible multiple stop itineraries, as verified in Table 6.6. This simple relaxation clearly keeps the relaxed problem very close to the original one and already decreases significantly the number of itineraries to be considered by the problem.

Table 6.6: Number of permitted itineraries for some different combinations of entry and exit airports configurations (from (a) to (d)) and inter-cluster restrictions (from (i) to (iv)).

	no relaxation	(d)(iv)	(c)(iii)	(b)(iii)	(b)(ii)	(b)(i)
direct routes F	**56**	**56**	**36**	**30**	**26**	**22**
one stop itineraries G	336	276	**124**	**92**	**68**	**48**
two stop itineraries H	1680	732	**266**	**174**	**118**	**78**
three stop itineraries I	6720	1296	384	224	**140**	**84**

Without the Cluster Problem Relaxation, all combinations between two airports for both ways were

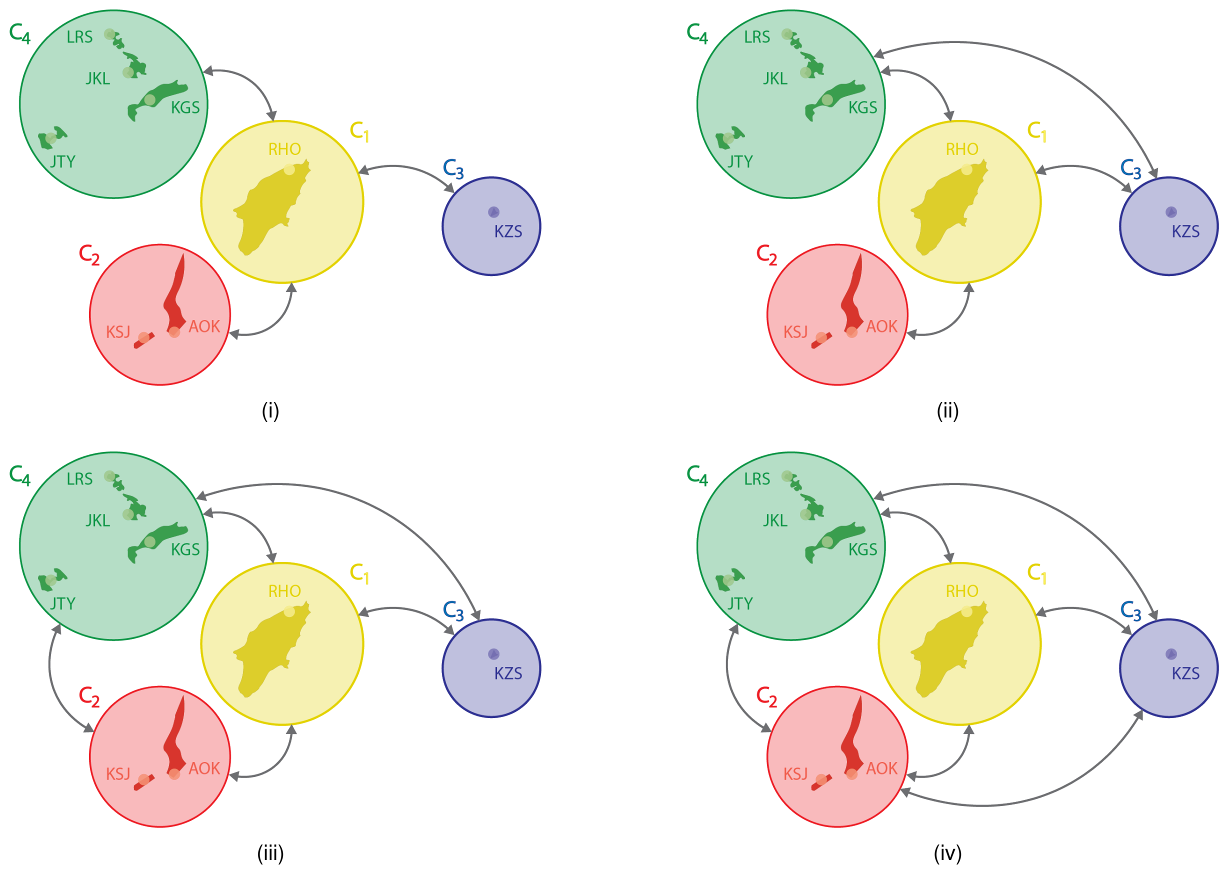


Figure 6.8: Different inter-cluster traveling restrictions.

considered as possible direct routes, resulting in 56 ($= 8 \times 7$). Consequently, there were 336 ($= 8 \times 7 \times 6$) one stop itineraries, 16140 ($= 8 \times 7 \times 6 \times 5$) two stop itineraries, and 6720 ($= 8 \times 7 \times 6 \times 5 \times 4$) three stop itineraries. Table 6.6 presents the number of permitted itineraries for some different combinations of entry and exit airports configurations (from (a) to (d)) and inter-cluster restrictions (from (i) to (iv)), with the itineraries needed to fulfill every link's demand highlighted with color and bold text. Note that having to include itineraries with more stops is not only less convenient for passengers, but also increases significantly the complexity of the optimization problem, as the number of itineraries to be considered by the problem increases, magnified by the fact that for each multiple stop itinerary a higher number of stops corresponds to a higher number of possibilities to consider (for example, for a specific $i \in I$ or $h \in H$, there are more possible combinations of $u_3(i, t, d, w_1, w_2, w_3, s)$ than of $u_2(h, t, d, w_1, w_2, s)$, with $t \in T$, w_1, w_2, and $w_3 \in W$, $s \in S$, because of the additional waiting time w_3 to be determined).

6.2.3 Partition of Passenger Demand

Since the case study involves extremely low demand routes, the idea of simply dividing the weekly passenger demand of each link by different days was readily discarded. For example, distributing the estimated 2 weekly low season passengers that want to travel from KSJ to JTY by two different days instead of allocating them to the same flight would most likely increase costs.

The alternative strategy found was to create two sets of passenger demand: *set 1* that only flies the most heavy demand routes, and *set 2* that flies all routes, including some passengers of the heavy demand routes. These two sets themselves could be distributed by more than one day.

The method to divide passenger demand by the two sets is schematized in Figure 6.9. Links with an estimated demand lower or equal to k are allocated to *set 2*. For links with an estimated demand higher than k, several different approaches were considered, some schematized in Figure 6.9: i) allocating proportionally those passengers, adjusting parameter α, ii) defining a constant number of passengers, c, to be allocated in *set 1* and allocate the remaining passengers to *set 2*, or even iii) allocating the minimum multiple of k to *set 1* so that only less than k passengers of that link remain on *set 2*.

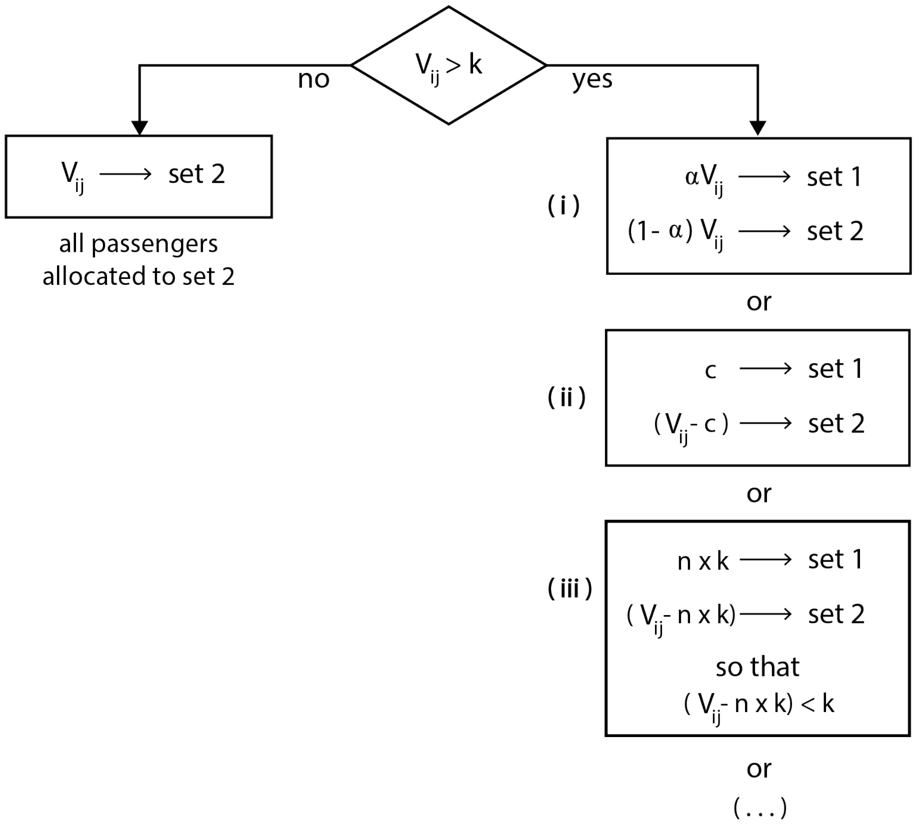


Figure 6.9: Possible approaches to passenger demand partition in sets.

Approach (iii) was selected, as it allowed the number of passengers allocated to set 1 to be easily manipulated. Constant k was set to be 8, as 14 of the 56 links (25%) have more than 8 weekly passengers (considering the median *Model 3* estimation), that already consists on a rather complex problem.

Chapter 7

Results

This chapter presents the results of applying the developed methods to the case study air network design. Firstly, the results from the Passenger Demand Prediction models are discussed in Section 7.1. Then, Section 7.2 presents the results from applying the Route and Flight Schedule Optimization program with different relaxation methods and discusses the most interesting ones.

7.1 Passenger Demand Prediction

A discussion on the goodness of fit (evaluated through the R^2 value) and a comparison, also taking the complexity into account (using the AIC), among the developed models, were presented throughout Section 5.2 and summarized in Section 5.2.7. In this section, *Models 3* and *7*, that consider the impact of population, touristic attractiveness, seasonality, and sea transport competition, are selected as the models worth of further analysis. Section 5.3 explains the limitations of the obtained models due to assumptions made regarding the available calibration data and introduces the necessity for the estimations to be normalized.

The distribution of the iterations estimated through the MC simulations and the selection of scenarios was also presented in Section 5.3, as they were important to explain the implementation of the methodological approach of the present work as a whole. Hereafter, these intermediate results will be discussed. Comments on the visible impact of seasonality on the results from *Models 3* and *7*, and then of the pandemic context on the results from *Model 7* will be presented, by using box-plot graphs.

Figure 7.1 allows the observation of the impact of seasonality by showing the observed values, represented by the black diamond shapes, and the scenarios estimated with *Model 3* and with *Model 7*, for $p_C = 1/3$. The colored dots represent the median and percentiles in accordance with Figures 5.6 and 5.7; the whiskers extend to the most extreme data points not considered outliers (i.e. between the 5th and 95th percentiles), and the outliers are plotted individually using black dots. Note that, for both *Model 3* and *7*, the most likely estimated scenarios of monthly total demand are higher than the observed for low season, and smaller for high season, leading to a smaller seasonal disparity than in the observed data.

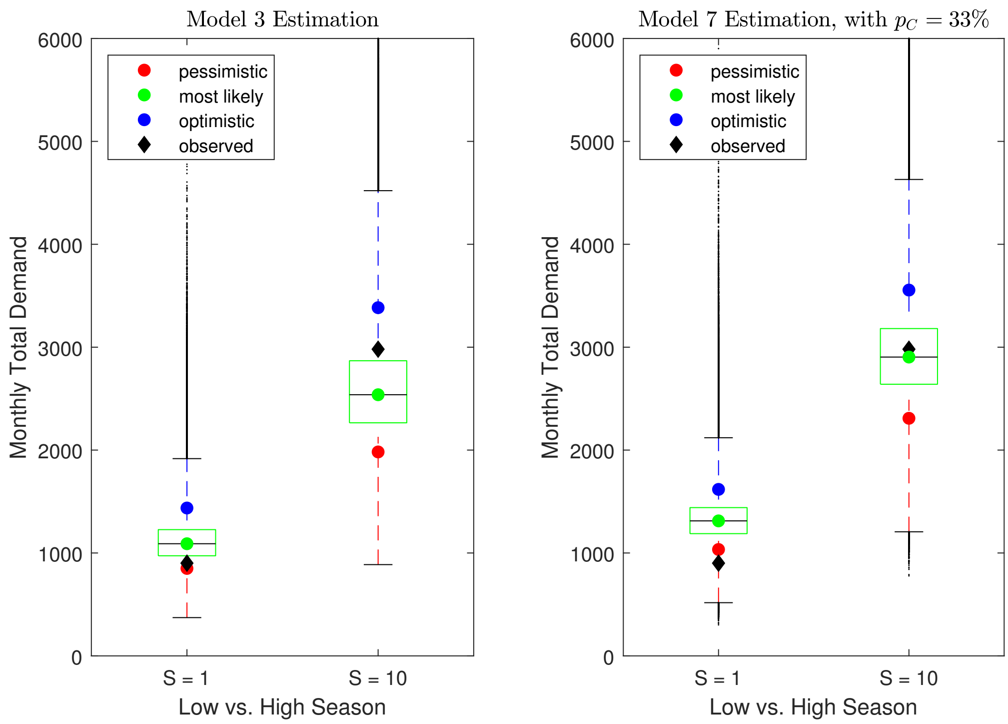


Figure 7.1: Box-plot of *Models 3* and *7* monthly total passenger demand iteration frequency results.

In Figure 5.7, one could observe that the pessimistic and most likely scenarios assume lower values when referring to the pandemic context ($C = 1$), as expected. However, that difference is not clear for the optimistic scenarios, marked with blue dots. A Levene's test was performed and showed a practically null p-value, excluding the null hypotheses that the population variances are equal. Variances for $C = 1$ are greater than for $C = 0$, as presented in Figure 7.2 (green box's height). This variance disparity results on closer optimistic scenarios for both outside and within a pandemic context, while the pessimistic scenarios show a much lower passenger demand prediction for the pandemic context. Even so, a steeper difference between outside and within the pandemic context for observed monthly total demand is verified (black diamond shapes), for both high and low seasons.

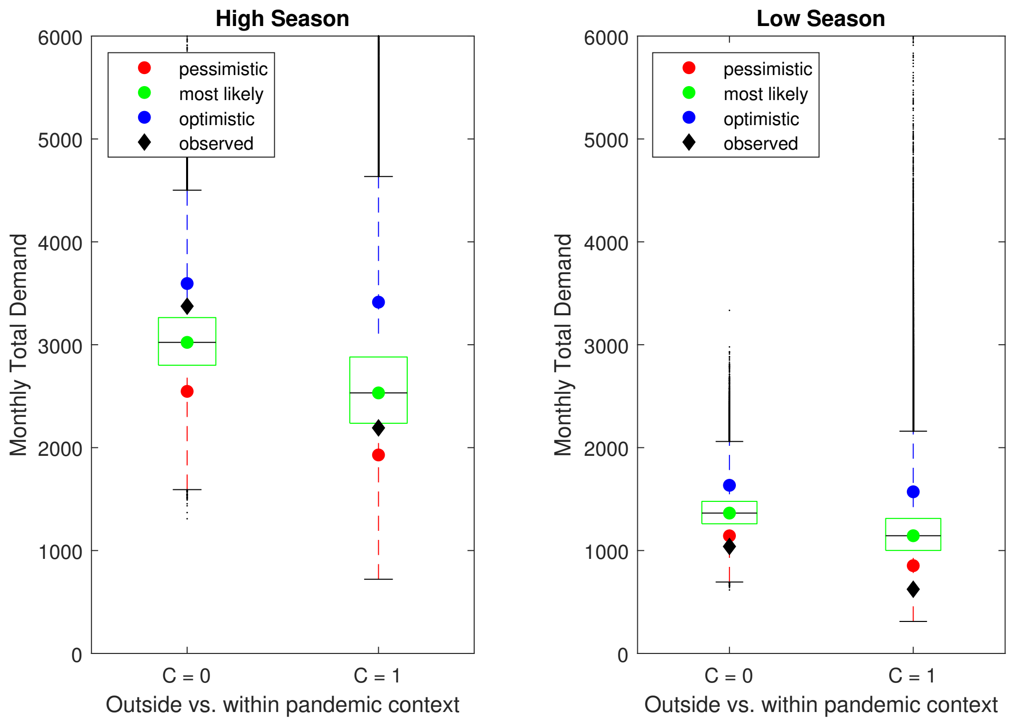


Figure 7.2: Box-plot of *Model 7* monthly total passenger demand iteration frequency results.

7.2 Route and Flight Schedule Optimization

This section presents and compares the most interesting results from the application of the Route and Flight Schedule Optimization program to the case study, using the passenger demand scenarios predicted with *Model 3* and using different combinations of operational time periods and relaxation techniques: cluster restriction with passenger demand partition in sets or not. These results were obtained on an Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz with 16 GB of RAM using the optimizer version 35.01.01 of the IBM's FICO Xpress software.

From the different combinations of entry and exit airports configurations (from (a) to (d)) and inter-cluster restrictions (from (i) to (iv)), (b)(iii) was considered to be a good compromise that reduced the complexity of the problem (as it only obliges multiple stop itineraries with up to two stops) while remaining fairly close to the original problem. This cluster restriction configuration was used for the results presented hereafter and summarized in Table 7.1. Figure 7.3 shows the evolution of the optimality gap while running the program using *Model 3* estimates considering two days with 12 operational hours each, without the demand partition relaxation. Figure 7.4 presents the best solution found, where purple stands for the two aircraft type 1 and orange for aircraft type 2, with an associated cost of 82,663 €. This solution is represented in Figure 7.3 by the last green square. At the time this solution was found (in 34,299.3s of computation time, approximately 9 hours and a half), the optimality gap assumed 66.24% and only improved to 65.90%, by the 24 hours of total computation time.

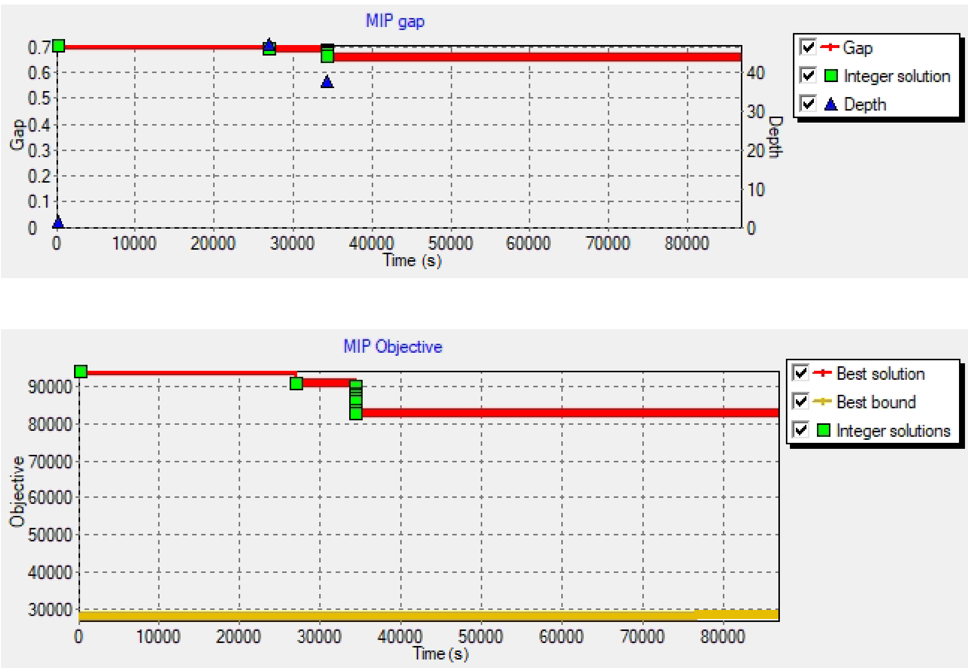


Figure 7.3: MIP search optimality gap, solutions depth, and objective evolution for problem with passenger demand estimated with *Model 3*, cluster relaxation (b)(iii), no passenger partition, $ND = 2$, $NT = 24$.

Figure 7.6 shows the results of dividing the passenger demand in two sets, according to Section 6.2.3, one set corresponding to each of the two days with 12 operational hours each (*set 1* allocated to day 1 and *set 2* to day 2). The solution for *set 1* corresponds to an 8.53% optimality gap, reached

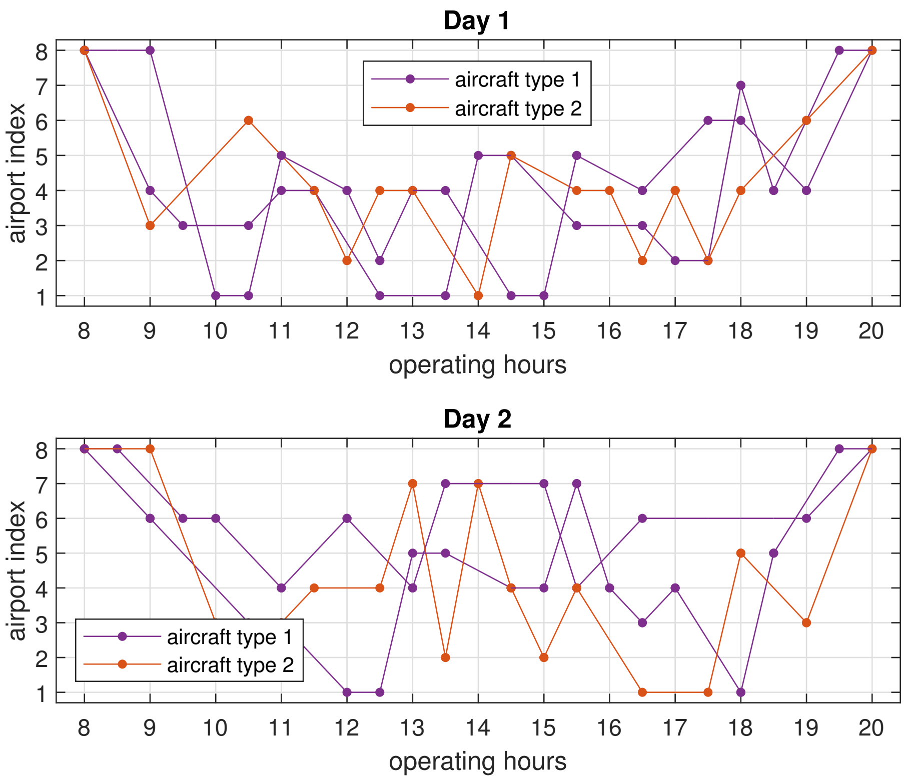


Figure 7.4: Solution for problem with passenger demand estimated with *Model 3*, cluster relaxation (b)(iii), no passenger partition, $ND = 2$, $NT = 24$.

after 1 hour and a half (5,429.8s), as presented in Figure 7.5. This relatively low computational time is attributable to the lower level of complexity of the problem, as it only handles passengers from 25% of the links. The solution for *set 2* corresponds to a 56.58% gap, after 16 hours (57908.3s), as also presented in Figure 7.5. Note that, if 0% optimality gaps were verified for both computations, it was expected that the cost obtained without further relaxing the problem through passenger partition would be higher than without dividing the passengers into sets, because the more the problem is relaxed, the less similar it stands to the original one. However, even with less computation time, the cost associated with this solution is actually lower (35,799 + 19,346 = 55,145 € $<$ 82,663 €), which shows the benefits of relaxing the optimization problem, as the original one cannot be handled within a reasonable computation time.

Another interesting result is the allocation of the whole passenger demand volume to one single operational day, with the same 12h. This result, presented in Figure 7.8, takes a 43.18% optimality gap, with 24 hours of computation time (decreasing insignificantly since the first hour and a half, as shown in Figure 7.7). 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 42,321 + 150$\times$3$\times$12 = 45,021 € (cost of three aircraft parked for 12 hours with a cost of 150€/h added to the cost of the 1 day of operations). 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, i.e. considering just the most likely scenario,

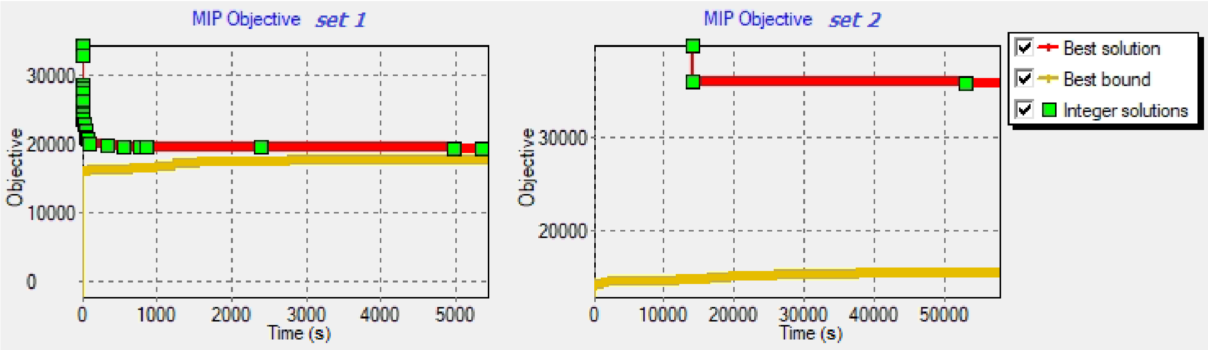


Figure 7.5: MIP search optimality gap, solutions depth, and objective evolution for problem with passenger demand estimated with *Model 3*, cluster relaxation (b)(iii) with passenger partition: two sets with $ND = 1$, $NT = 24$, each.

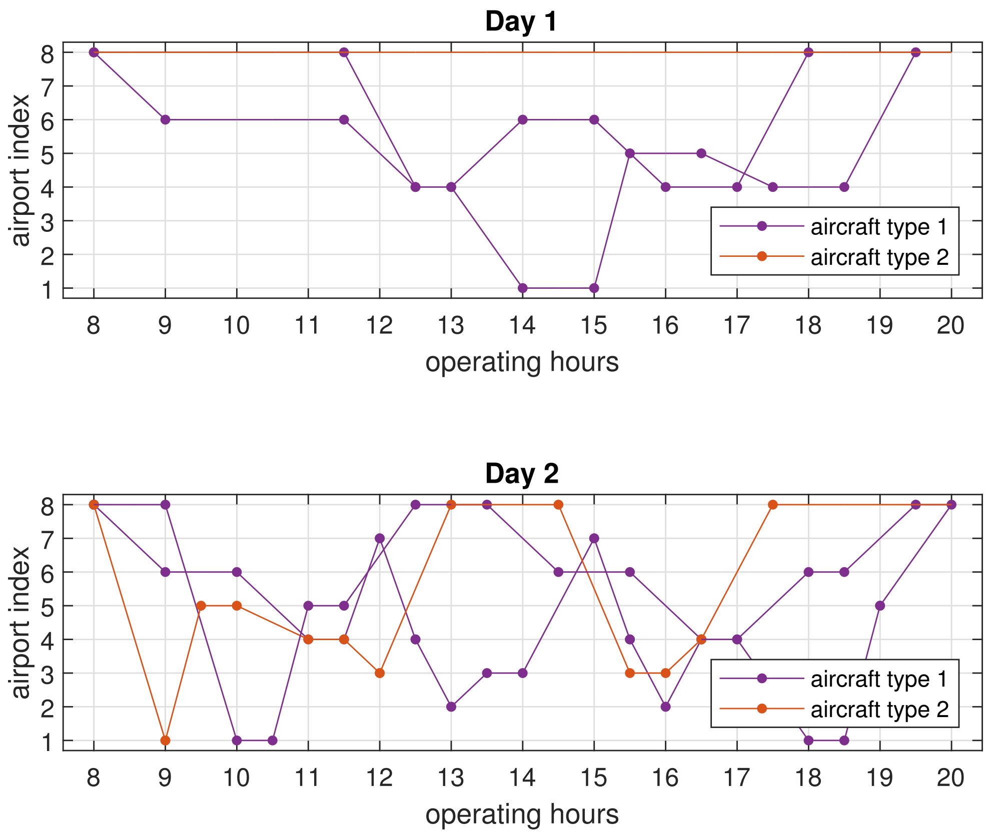
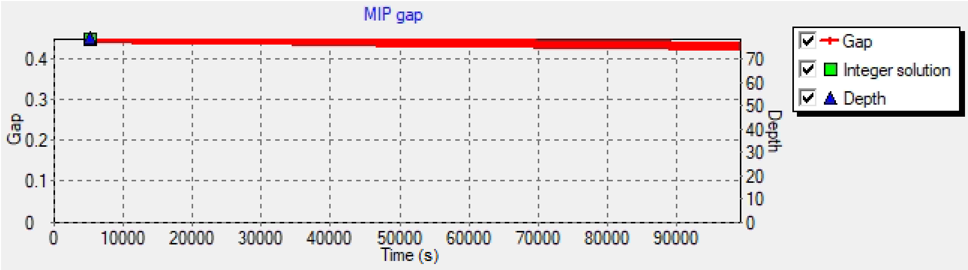


Figure 7.6: Solution for problem with passenger demand estimated with *Model 3*, cluster relaxation (b)(iii), with passenger partition: two sets with $ND = 1$, $NT = 24$, each.

for $ND = 1$, $NT = 24$ led to the solution in Figure 7.9. This solution corresponds to a 32.50% optimality gap, lower than the 43.18% from the 3 scenario approach, and was reached 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 7.2. 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. The cost for the



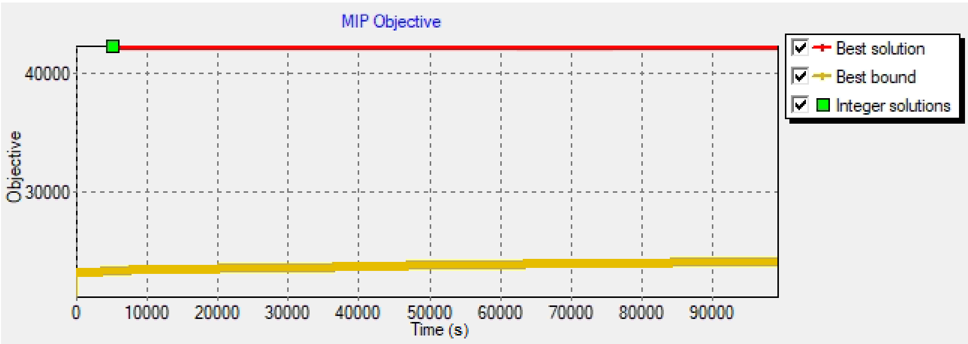


Figure 7.7: MIP search optimality gap, solutions depth, and objective evolution for problem with passenger demand estimated with *Model 3*, cluster relaxation (b)(iii), no passenger partition, $ND = 1$, $NT = 24$.

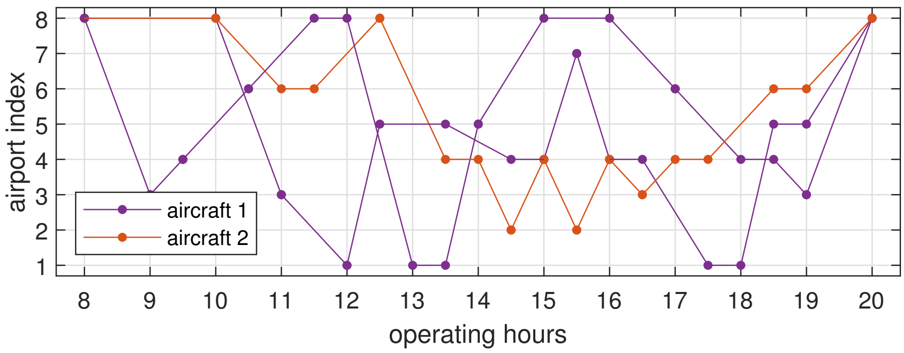


Figure 7.8: Solution for problem with passenger demand estimated with *Model 3*, cluster relaxation (b)(iii), no passenger partition, $ND = 1$, $NT = 24$.

pessimistic scenario was also lower for the deterministic approach. 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.

The major conclusion one can draw from these results is that the relaxation of the optimization problem is essential to obtain lower network costs. Although the best bound cost is higher for a relaxed problem, the best solution decreases much faster, allowing to obtain better results within a reasonable computation time. It also stood clear that reducing the operational time for the program to consider drastically reduces the complexity of the problem, leading to better results.

It was shown that, although considering only the most likely scenario reduces the complexity of the problem and leads to a lower network cost for that scenario, this solution is not capable of serving an optimistic passenger demand scenario. Therefore, the Route and Flight Schedule Optimization program

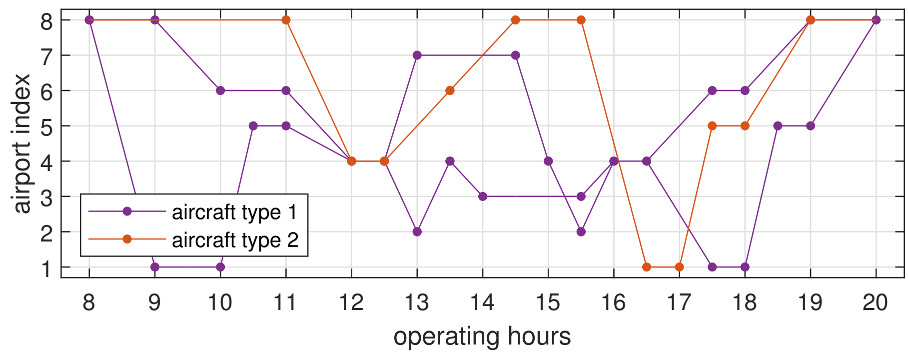


Figure 7.9: Solution for problem with passenger demand estimated with *Model 3*, cluster relaxation (b)(iii), no passenger partition, $ND = 1$, $NT = 24$, considering just the most likely scenario.

Table 7.1: Optimization results, applying *Model 3*, with cluster restriction b(iii).

	No pass. partition	Partition in 2 *sets*		No pass. partition	Just the most likely scenario
		set 1	*set 2*		
	$ND = 2$ $NT = 24$	$ND = 1$ $NT = 24$	$ND = 1$ $NT = 24$	$ND = 1$ $NT = 24$	$ND = 1$ $NT = 24$
computation time (s)	86400.0	5429.8	57908.3	99113.2	60412.3
optimality gap	65.90%	8.53%	56.58%	43.18%	32.50%
estimated cost (€)	82663	19346	35799	42321	33127

Table 7.2: Calculation of estimated costs.

Costs		Three scen. opt.	Most likely scen. opt.
Operating flights (€)	O_1	36799	27787
Parking aircraft (€)	O_2	300	450
Passengers time (€)	Pessimistic ($s = 1$)	4130	3975
	Most likely ($s = 2$)	5055	4890
assuming $p(s) = 1,\ s = 1, 2, 3$	Optimistic ($s = 3$)	6480	unfeasible
Expected value (€)	$\sum_{k=1}^{9} O_k$	42321	-

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

Chapter 8

Conclusions

In this final chapter, a summary of the most relevant achievements accomplished by the present work is presented. The limitations of the developed methods are also enumerated, followed by suggestions for future work to overcome those limitations and further develop this topic.

8.1 Achievements

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.

8.2 Limitations

There is an important limitation regarding the demand prediction models observed, thoroughly discussed in Section 5.3, due to the available data-set. This limitation hinders the estimated volume of passenger demand for the entire network in this case study, as explained before, but does not forfeit the applicability of these models to more suitable data-sets. Additionally, the ferry competition factor (Section 5.2.4) should be more carefully studied, as the offered sea transport service strongly varies throughout the year depending on the season.

The major limitation of the Route and Flight Schedule Optimization program developed stands with its tractability issues. The optimality gaps associated with the network solutions presented, computed within up to 25 hours, were far from ideal. Additionally, only results for considering low season demand were present, and it is known that considering high season demand, with the same fleet, would oblige larger operational time periods, increasing even more the complexity of the problem. If these issues were to be surpassed, a higher number of different passenger demand scenarios could be given to the problem to consider simultaneously, which would improve the robustness of this tool.

8.3 Future Work

A partnership with Greek authorities to facilitate the acquisition of more data about the different islands would be extremely beneficial, enhancing the data-set with more explanatory variables and allowing real distinction between islands within the same NUTS 3 regions. The results from using that more suitable data-set to predict passenger demand would allow a real application of the knowledge depicted in the present work to the existing network. As there are a total of 44 airports in Greece, there is an opportunity to extend this study to more islands, not only so that the whole network could be further optimized using the decision support framework considered in the present work, but also so that the results of the passenger demand prediction model could give some insight to redefine the airports to be included in the PSO network before applying an SFSFA optimization tool.

As previously stated, ferry competition should be more carefully explored. It could also be beneficial to develop an integrated decision support model, for both sea transport and air transport subsidization simultaneously. Although the tender would be separate for ferry companies and airlines, the minimum service to grant to users should be distributed by these two modes of transportation, implying the decision on the requirements for air transport PSO and sea transport PSO to be embedded in one fully integrated process.

With respect to the Route and Flight Schedule Optimization program, it would be of great benefit to develop meta-heuristic methods to make the solution search faster. Another interesting enhancement would be to consider a feedback loop within the optimization problem to account for the effect of frequency, flight pricing and other network characteristics on passenger demand. There is the opportunity of including more features, namely regarding fleet assignment. Upon the social cost computation, it would be useful to consider the convenience for passengers to reach their destination and come back to their origin island within the same day.

Regarding the decision support method suggested as a whole, the present work was based on the idea that if a limited amount of requirements was defined in the PSO tender, giving more freedom to the airline entrants, the airlines would optimize the network taking only their own interests into account, hindering the application of an optimal solution for all stakeholders. However, defining such narrow requirements is not in line with EU guidelines and Kinene [20] points out that airlines claim that strict PSO requirements are a major factor that discourages their participation in the tendering process, causing an alarming decrease of airline entrants. Furthermore, the idea that the responsible authorities should be able to perform a deep study of every region and develop a route and flight schedule solution that is optimal for the next four to five years is not reasonable. Firstly, this type of research is expensive and time-consuming. Secondly, even a robust solution should be regularly adapted to the circumstances, meaning that there should be a constant assessment of the regions' connectivity to feedback and adjust the air transportation network, which is not compatible with the current PSO tendering process. If the government was to do an exhaustive plan of operation and reiterate it frequently, it might as well run its own public airline, because at that point, the subsidized airline would only be providing the aircraft and crew, with no decision power on the operational plan. Henceforth, a decision model approach similar to Kinene [20], that leaves freedom to the subsidized airline and takes a decision based on the prediction of its behavior, could be pursued.

Bibliography

[1] A. Smyth, G. Christodoulou, N. Dennis, A.-A. Marwan, and J. Campbell. Is air transport a necessity for social inclusion and economic development? *Journal of Air Transport Management*, 22:53–59, 2012.

[2] F. Zhang and D. J. Graham. Air transport and economic growth: a review of the impact mechanism and causal relationships. *Transport Reviews*, 40(4):506–528, 2020.

[3] Regulation (ec) no 1008/2008 of the European Parliament and of the Council of 24 september 2008 on common rules for the operation of air services in the Community. *Official Journal of the European Union*, 31.10.2008.

[4] European Comission Notice on Interpretative guidelines on Regulation (EC) No 1008/2008 of the European Parliament and of the Council — Public Service Obligations (PSO). *Official Journal of the European Union*, 17.6.2017.

[5] 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.

[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, 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.

[8] R. Devoto, C. Farci, and F. Lilliu. Analysis and forecast of air transport demand in Sardinia's airports as a function of tourism variables. *WIT Transactions on The Built Environment*, 60, 2002.

[9] J. Chevallier, B. Chèze, and P. Gastineau. Forecasting world and regional air traffic in the mid-term (2025): An econometric analysis of air traffic determinants using dynamic panel-data models. 2011.

[10] T. Tsekeris. Dynamic analysis of air travel demand in competitive island markets. *Journal of Air Transport Management*, 15(6):267–273, 2009.

[11] A. Nõmmik and S. Kukemelk. Developing gravity model for airline regional route modelling. *Aviation*, 20(1):32–37, 2016.

[12] Z. Wadud. Modeling and forecasting passenger demand for a new domestic airport with limited data. *Transportation research record*, 2214(1):59–68, 2011.

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

[14] R. Doganis. Traffic forecasting and the gravity model. *Flight International*, 29:547–549, 1966.

[15] F. Kopsch. A demand model for domestic air travel in Sweden. *Journal of Air Transport Management*, 20:46–48, 2012.

[16] E. Sambracos and K. Rigas. Passenger reactions to market deregulation: First results from the experience of the Greek islands market. *Journal of Air Transport Management*, 13(2):61–66, 2007.

[17] H. Ito and D. Lee. Assessing the impact of the September 11 terrorist attacks on US airline demand. *Journal of Economics and Business*, 57(1):75–95, 2005.

[18] S. Dožić and M. Kalić. Three-stage airline fleet planning model. *Journal of air transport management*, 46:30–39, 2015.

[19] C. Iliopoulou, K. Kepaptsoglou, and M. G. Karlaftis. Route planning for a seaplane service: The case of the greek islands. *Computers & Operations Research*, 59:66–77, 2015.

[20] A. Kinene. *Decision Models for the Procurement of Subsidised Air Services*. PhD thesis, Linköping University Electronic Press, 2021.

[21] A. Kinene, T. A. Granberg, V. Polishchuk, and C. Rydergren. Decision support for an optimal choice of subsidised routes in air transportation. *Journal of Air Transport Management*, 82:101724, 2020.

[22] A. P. Antunes, M. G. Santos, J. P. Pita, and A. G. Menezes. Study on the evolution of the air transport network of the azores. *Transportation Research Part A: Policy and Practice*, 118:837–851, 2018.

[23] A. Alin. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):370–374, 2010.

[24] D. M. Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

[25] L. Mendo and J. M. Hernando. A simple sequential stopping rule for monte carlo simulation. *IEEE Transactions on Communications*, 54(2):231–241, 2006.

[26] List of public service obligations as of 18.9.2019. *European Comission Website*, accessed on 22.09.2021. URL `https://ec.europa.eu/transport/modes/air/internal-market/pso_en`.

[27] Research for TRAN Committee - Transport and Tourism in Greece, 16.4.2018. *European Parliament Website*, accessed on 24.09.2021. URL https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_BRI(2018)617469.

[28] Tourism statistics at regional level. *Eurostat Statistics Explained*, accessed on 24.09.2021. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism_statistics_at_regional_level#Number_of_nights_spent.

[29] Properties inscribed on the world heritage list. *UNESCO Website*, accessed on 24.09.2021. URL https://whc.unesco.org/en/statesparties/gr.

[30] Nights spent at tourist accommodation establishments by nuts 2 regions. *Eurostat Data Browser*, accessed on 24.09.2021. URL https://ec.europa.eu/eurostat/databrowser/view/tgs00111/default/table?lang=en.

[31] 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. URL https://ec.europa.eu/eurostat/databrowser/view/tin00172/default/map?lang=en.

[32] Civil Aviation Authority Organization and Development Division Statistics Section. Air traffic statistics. accessed on 19.05.2021. URL http://www.ypa.gr/en/profile/statistics/.

[33] S. Birolini, A. P. Antunes, M. Cattaneo, P. Malighetti, and S. Paleari. Integrated flight scheduling and fleet assignment with improved supply-demand interactions. *Transportation Research Part B: Methodological*, 149:162–180, 2021.