ROBUSTNESS AND RECOVERABILITY IN TRANSPORT LOGISTICS

LUIS CADARSO MORGÁ
B.Sc. AEROSPACE ENGINEERING (2009)
M.Sc. AEROSPACE ENGINEERING (2011)

DIRECTOR
ÁNGEL MARÍN GRACIA
PROFESSOR OF APPLIED MATHEMATICS IN AEROSPACE ENGINEERING

2013
Abstract

Transport is the foundation of any economy: it boosts economic growth, creates wealth, enhances trade, geographical accessibility and the mobility of people. Transport is also a key ingredient for a high quality of life, making places accessible and bringing people together. The future prosperity of our world will depend on the ability of all of its regions to remain fully and competitively integrated in the world economy. Efficient transport is vital in making this happen.

Operations research can help in efficiently planning the design and operating transport systems. Planning and operational processes are fields that are rich in combinatorial optimization problems. These problems can be analyzed and solved through the application of mathematical models and optimization techniques, which may lead to an improvement in the performance of the transport system, as well as to a reduction in the time required for solving these problems. The latter aspect is important, because it increases the flexibility of the system: the system can adapt in a faster way to changes in the environment (i.e.: weather conditions, crew illness, failures, etc.). These disturbing changes (called disruptions) often enforce the schedule to be adapted. The direct consequences are delays and cancellations, implying many schedule adjustments and huge costs. Consequently, robust schedules and recovery plans must be developed in order to fight against disruptions.

This dissertation makes contributions to two different fields: rail and air applications. Robust planning and recovery methods are presented.

In the field of railway transport we develop several mathematical models which answer to RENFE’s (the major railway operator in Spain) needs:

1. We study the rolling stock assignment problem: here, we introduce some robust aspects in order to ameliorate some operations which are likely to fail. Once the rolling stock assignment is known, we propose a robust routing model which aims at identifying the train units’ sequences while minimizing the expected delays and human resources needed to perform the sequences.
2. It is widely accepted that the sequential solving approach produces solutions that are not global optima. Therefore, we develop an integrated and robust model to determine the train schedule and rolling stock assignment. We also propose an integrated model to study the rolling stock circulations. Circulations are determined by the rolling stock assignment and routing of the train units.

3. Although our aim is to develop robust plans, disruptions will be likely to occur and recovery methods will be needed. Therefore, we propose a recovery method which aims to recover the train schedule and rolling stock assignment in an integrated fashion all while considering the passenger demand.

In the field of air transport we develop several mathematical models which answer to IBERIA’s (the major airline in Spain) needs:

1. We look at the airline-scheduling problem and develop an integrated approach that optimizes schedule design, fleet assignment and passenger use so as to reduce costs and create fewer incompatibilities between decisions. Robust itineraries are created to ameliorate misconnected passengers.

2. Air transport operators are continuously facing competition from other air operators and different modes of transport (e.g., High Speed Rail). Consequently, airline profitability is critically influenced by the airline’s ability to estimate passenger demands and construct profitable flight schedules. We consider multi-modal competition including airline and rail, and develop a new approach that estimates the demand associated with a given schedule; and generates airline schedules and fleet assignments using an integrated schedule design and fleet assignment optimization model that captures the impacts of schedule decisions on passenger demand.

**Key words:** Railway Scheduling, Airline Scheduling, Integration, Robustness, Competition, Recoverability, Passenger Demand, Schedule Design, Rapid Transit, Rolling Stock, Routing, Fleet assignment.
Resumen

El transporte es la base de cualquier economía: impulsa el crecimiento económico, crea riqueza, aumenta el comercio, la accesibilidad geográfica y la movilidad de las personas. El transporte es también un ingrediente clave para una alta calidad de vida, haciendo los lugares accesibles y uniendo a las personas. La prosperidad futura de nuestro mundo dependerá de la capacidad de todas sus regiones para permanecer completamente integradas de manera competitiva en la economía mundial. Un transporte eficiente es vital para hacer que esto suceda.

La investigación operativa puede ayudar eficazmente en la planificación del diseño y gestión del transporte. Los procesos de planificación y gestión son campos que son ricos en problemas de optimización combinatoria. Estos problemas pueden ser analizados y resueltos por medio de la aplicación de modelos matemáticos y técnicas de optimización, que pueden conducir a una mejora en el rendimiento del sistema, así como a una reducción en el tiempo requerido para resolver estos problemas. Este último aspecto es importante, ya que aumenta la flexibilidad del sistema: el sistema se puede adaptar de forma más rápida a los cambios en el medio ambiente (por ejemplo, las condiciones climáticas, las enfermedades de la tripulación, fallos, etc.). Estos cambios anormales (llamados interrupciones o rupturas) a menudo hacen cambiar la programación del sistema para adaptarse a las nuevas circunstancias. Las consecuencias directas son retrasos y cancelaciones, lo que implica muchos ajustes en los programas y costes enormes. Por lo tanto, es necesario desarrollar tanto programas robustos como planes de recuperación con el fin de luchar contra las interrupciones o rupturas del sistema.

Esta tesis contribuye a dos campos diferentes: las aplicaciones ferroviarias y aéreas. Se presentan métodos de planificación robusta y métodos de recuperación.

En el ámbito del transporte ferroviario, desarrollamos varios modelos matemáticos que responden a las necesidades de RENFE (la mayor compañía ferroviaria en España):

1. Se estudia el problema de asignación del material rodante: introducimos robustez en el sistema con el fin de mejorar algunas operaciones que son propensas a fallar.
Una vez que la asignación de material rodante es conocida, proponemos un modelo de enrutamiento robusto que tiene por objeto la identificación de las secuencias del material rodante y la reducción al mínimo los retrasos previstos y los recursos humanos necesarios.

2. Es un hecho ampliamente aceptado que el enfoque secuencial para resolver la planificación de los sistemas de transporte produce soluciones que no son óptimos globales. Por lo tanto, desarrollamos un modelo integrado y robusto para determinar el horario de los servicios y la asignación de material rodante. También proponemos un modelo integrado para estudiar la circulación del material rodante. Las circulaciones están determinadas por la asignación de material rodante y el enrutamiento del mismo.

3. Aunque nuestro objetivo es desarrollar planes robustos, las interrupciones son probables y por lo tanto es necesario desarrollar planes de recuperación. Proponemos un método de recuperación que tiene como objetivo recuperar los horarios de los servicios y la asignación de material rodante de forma integrada, teniendo en cuenta la demanda de pasajeros.

En el ámbito del transporte aéreo, desarrollamos varios modelos matemáticos que responden a las necesidades de IBERIA (la principal línea aérea de España):

1. Nos fijamos en el problema de la programación de las aerolíneas y desarrollamos un enfoque integrado que optimiza el diseño del programa, la asignación de la flota y el uso del pasajero con el fin de reducir costes y crear menos incompatibilidades entre las diferentes decisiones. Creamos itinerarios robustos para evitar las pérdidas de conexión de los pasajeros.

2. Los operadores de transporte aéreo se enfrentan continuamente a la competencia de otros operadores aéreos y los diferentes modos de transporte (por ejemplo, el tren de alta velocidad). En consecuencia, la rentabilidad de las aerolíneas está gravemente influenciada por la capacidad de la aerolínea para calcular las demandas de los pasajeros y construir programas de vuelo rentables. Consideramos la competencia multimodal incluyendo las compañías aéreas y el ferrocarril, y desarrollamos un nuevo enfoque que estima la demanda asociada a un programa determinado, y genera tanto los horarios como la asignación de la flota mediante un modelo de optimización integrado que captura los impactos de las decisiones de la planificación en la demanda de pasajeros.
Palabras clave: Planificación en Ferrocarril, Planificación en Aerolíneas, Integración, Robustez, Competencia, Recuperabilidad, Demanda de Pasajeros, Diseño de la Programación, Tránsito Rápido, Material Rodante, Enrutamiento, Asignación de la flota.
Acknowledgments

First, I would like to thank my research advisor, Professor Ángel Marín. This thesis would have not been possible without his support. He was always available to discuss the research related topics.

I am fortunate to have worked with Gábor Maróti and Leo Kroon at the Rotterdam School of Management. They kindly agreed to host me during a research visit. Their vast knowledge about railways has contributed to actively improve this thesis’s contents. I would also like to thank Lars Kjær Nielsen for being available during the first phase of this research.

I am grateful to Cynthia Barnhart for agreeing to host me during a research visit at the Massachusetts Institute of Technology. I would also like to thank Vikrant Vaze and Virot Ta Chiraphadhanakul for their helpful comments and suggestions on the research driven at the Massachusetts Institute of Technology. All of them have been available to constructively discuss about the research during and after my visit. Undoubtedly, their great knowledge about airlines and continued support have made possible to satisfactorily conclude some of the research in this thesis.

I also want to thank to Leo Kroon and Bernard Gendron for the time they have spent reading this thesis, and Ángel Luis Matesanz from RENFE and Lorenzo del Castillo from IBERIA for their support in the data collection.

I have to thank the Department of Applied Mathematics and Statistics from the Higher Technical School of Aeronautical Engineering, the Higher Technical School of Aeronautical Engineering from the Technical University of Madrid, the Technical University of Madrid and the Spanish government (Project grants: CEDEX, Ministerio de Fomento, PT-2007-003-08CCPP; Ministerio de Ciencia e Innovación, TRA2008-06782-C02-01; and Ministerio de Economía y Competitividad, TRA2011-27791-C03-01) for their support.

I also want to thank my friends who have suggested me minor and/or major changes to this dissertation.

Finally, I would like to thank my family. Nothing would have been possible without
the constant love, support and encouragement from my parents. I am forever indebted to them for all the sacrifices they have made for me throughout my life.
## Contents

Abstract iv

Resumen viii

Acknowledgments xiii

### 1 INTRODUCTION

#### 1 OPERATIONS RESEARCH IN TRANSPORT ENGINEERING

1.1 Introduction .................................................. 5
1.2 Planning in Transport ........................................... 6
1.3 Recoverability and Robustness .................................. 9

#### 2 THESIS OUTLINE

2.1 Chapter 3 Thesis Contributions ................................. 11
2.2 Chapter 4 Introduction ........................................... 11
2.3 Chapter 5 Robust Rolling Stock in Rapid Transit Networks .......... 11
2.4 Chapter 6 Robust Routing of Rapid Transit Rolling Stock .......... 12
2.5 Chapter 7 Integration of Timetable Planning and Rolling Stock in Rapid Transit Networks ................. 13
2.6 Chapter 8 Improving Robustness of Rolling Stock Circulations in Rapid Transit Networks ...................... 13
2.7 Chapter 9 Recovery of Disruptions in Rapid Transit Networks ......... 14
2.8 Chapter 10 Introduction ........................................... 14
2.9 Chapter 11 Robust Passenger Oriented Timetable and Fleet Assignment Integration in Airline Planning ................. 14
5 ROBUST ROLLING STOCK IN RAPID TRANSIT NETWORKS

5.1 Introduction .................................................. 37
5.2 Literature Review ........................................... 39
5.3 The Rolling Stock Problem in Rapid Transit Networks ............. 41
5.4 The Rolling Stock Model ...................................... 43
  5.4.1 Objective Function .......................................... 47
  5.4.2 Train Service Constraints ................................. 48
  5.4.3 Demand Constraints ........................................ 48
  5.4.4 Material Constraints ....................................... 48
  5.4.5 Shunting Constraints ........................................ 49
  5.4.6 Variable Domain .......................................... 50
5.5 Computational Experiments .................................. 51
  5.5.1 Study Case 1: Line C5 .................................... 51
  5.5.2 Study Case 2: Lines C3-C4 ................................. 55
5.6 Summary ..................................................... 57

6 ROBUST ROUTING OF RAPID TRANSIT ROLLING STOCK  

6.1 Introduction .................................................. 59
6.2 Literature Review ........................................... 60
6.3 Train Routing Problem in Rapid Transit Networks ..................... 61
6.4 The Rapid Transit Routing Model ................................ 65
6.5 Case-Study: Madrid Suburban Rail Network ......................... 67
  6.5.1 Line C5: Computational Results .............................. 68
  6.5.2 Lines C3 and C4: Computational Results ...................... 72
6.6 Summary ..................................................... 73

7 INTEGRATION OF TIMETABLE PLANNING AND ROLLING STOCK ASSIGNMENT IN RAPID TRANSIT NETWORKS

7.1 Introduction .................................................. 75
7.2 Literature Review ........................................... 77
7.3 Timetable Planning and Rolling Stock Problem in Rapid Transit Networks . 79
7.4 The Timetable Planning and Rolling Stock Model .................... 81
  7.4.1 Sets .................................................. 82
  7.4.2 Cost parameters ......................................... 83
  7.4.3 Other parameters ......................................... 83
  7.4.4 Variables .............................................. 84

xvii
List of Figures

4.1 RENFE Cercanías Madrid Rapid Transit Network ........................................ 32
5.1 Line C5 ........................................................................................................ 52
5.2 Line C3 ........................................................................................................ 55
5.3 Line C4 ........................................................................................................ 55
6.1 Sequence with three operations .................................................................... 63
6.2 Crew requirement penalization ...................................................................... 65
6.3 Graphic representation of feasible solution 1 for line C5 ............................... 71
6.4 Graphic representation of feasible solution 2 for line C5 ............................... 71
6.5 Graphic representation of the optimal RRTRM solution for line C5 ............... 72
7.1 Different RT&RS solutions for Line C5 obtained by varying $\tau(k)$ value ...... 92
7.2 Passengers in excess for the arcs with the greatest passenger flow in Line C5 93
7.3 Different RT&RS solutions for Lines C3&C4 obtained by varying $\tau(k)$ value 93
8.1 Line C5: number of operations connections for each propagated delay in the integrated approach (left side) and in the sequential approach (right side) 115
8.2 Line C3&C4: number of operations connections for each propagated delay in the integrated approach (left side) and in the sequential approach (right side) 117
9.1 Network topology while the blockage is active ............................................. 136
9.2 Passenger demand in the solution $P\&O$: number of requested trips (higher curve) and number of denied trips (lower curve) ........................................ 139
9.3 Passenger demand in the solution $Operator$: number of requested trips (higher curve) and number of denied trips (lower curve) ........................... 140
9.4 Passenger demand in the first and second iterations for the combination of lines C5 and C41/C42 ................................................................. 142
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.5</td>
<td>Passenger demand in the first and second iterations for the combination of line C5 and the Metro network</td>
<td>142</td>
</tr>
<tr>
<td>9.6</td>
<td>Passenger demand in the first and second iterations for the line C5</td>
<td>143</td>
</tr>
<tr>
<td>10.1</td>
<td>IBERIA’s national network</td>
<td>160</td>
</tr>
<tr>
<td>11.1</td>
<td>Available itineraries for a determined market</td>
<td>170</td>
</tr>
<tr>
<td>11.2</td>
<td>Available itineraries for passengers in market 309 in the non-robust and robust solutions</td>
<td>178</td>
</tr>
<tr>
<td>12.1</td>
<td>Comparison of the frequencies per fleet type given by the model with the airline’s actual schedule for $\rho = 1$</td>
<td>202</td>
</tr>
<tr>
<td>12.2</td>
<td>Comparison of the seats flown of the model with the airline’s actual schedule for $\rho = 1$</td>
<td>202</td>
</tr>
<tr>
<td>12.3</td>
<td>Percentage difference (with respect to actual numbers) in the air market share in the markets where HSR currently operates</td>
<td>204</td>
</tr>
<tr>
<td>12.4</td>
<td>Percentage difference in the number of transported passengers and in profit with respect to IBERIA’s actual values</td>
<td>204</td>
</tr>
<tr>
<td>12.5</td>
<td>Percentage difference in the fleet utilization with respect to IBERIA’s actual values</td>
<td>205</td>
</tr>
<tr>
<td>12.6</td>
<td>Model’s response to HSR in origin-destination pairs Madrid-Vigo, Madrid-La Coruña and Madrid-Santiago</td>
<td>209</td>
</tr>
<tr>
<td>12.7</td>
<td>Average load factor predicted by the model in response to HSR entry with varying frequencies in the origin-destination pairs: Madrid-Vigo and Vigo-Madrid</td>
<td>209</td>
</tr>
<tr>
<td>12.8</td>
<td>Model’s predicted market share in response to HSR in origin-destination pair Madrid-La Coruña and La Coruña-Madrid</td>
<td>210</td>
</tr>
<tr>
<td>12.9</td>
<td>Model’s predicted total profit variation in response to HSR entry for the base price and reductions of 10%, 20%, 30% and 40% in ticket prices</td>
<td>211</td>
</tr>
</tbody>
</table>
List of Tables

5.1 Train unit capacity and length .............................................. 52
5.2 Line C5 RS model: the number of variables, constraints and non-zeros . 52
5.3 Line C5 RS model solutions ...................................................... 54
5.4 Line C5 RS current solution operated by RENFE .......................... 54
5.5 Comparing non-robust and robust solutions for line C5 .................. 55
5.6 Train unit capacity and length ...................................................... 56
5.7 Lines C3-C4 RS model: the number of variables, constraints and non-zeros 56
5.8 Lines C3-C4 RS model solutions ................................................. 57
5.9 Comparing the non-robust and robust solutions for lines C3-C4 .......... 57
5.10 Current RS solution operated by RENFE for lines C3 and C4 .......... 57
6.1 Line C5 RRTRM model: number of variables, constraints and non-zeros . 68
6.2 Sequence given by RRTRM for train unit 34 .............................. 69
6.3 Sequence given by RRTRM for train unit 55 ................................ 69
6.4 Line C5: expected delays in minutes ......................................... 70
6.5 Line C3-C4 RRTRM model: number of variables, constraints and non-zeros 72
6.6 Lines C3-C4: expected delays in minutes .................................... 72
7.1 Train lines in Line C5 ............................................................... 90
7.2 Train lines in Lines C3&C4 .......................................................... 90
7.3 Line C5 RT&RS model: the number of variables, constraints and non-zeros 91
7.4 RT&RS model solution compared to RS with fixed timetable solution and current RENFE solution ......................................................... 94
7.5 Line C5 RT&RS model with and without maintenance costs comparison ........................................ 95
7.6 Comparing non-robust and robust solutions ..................................... 96
7.7 Comparing robust and swapping robustness solutions for line C5 ........ 97
8.1 Line C5 Benders based heuristic solution at each iteration ............... 114
8.2 Line C5: Benders heuristic solution, sequential (RS & TR) solution and current solution .......................................................... 116
8.3 Lines C3&c4 Benders based heuristic solution at each iteration .......... 116
8.4 Lines C3&c4 heuristic solution vs. current solution ............................. 117
9.1 Recovery solutions .............................................................................. 139
9.2 Frequency values in Line C5 between VIAL and OR .............................. 140
9.3 Recovery and original planning ............................................................. 141
11.1 IRASM: the number of variables, constraints and non-zeros ............... 176
11.2 Passengers comparisons between the non-robust and robust approaches ................................................................................. 177
11.3 Costs comparisons between the non-robust and robust approaches ........ 178
11.4 Fleet comparisons between the non-robust and robust approaches ....... 179
11.5 Fleet comparisons between the robust approaches with a minimum average block hour fleet utilization of 7 and 8 .................... 179
11.6 Passengers comparisons between the robust approaches with a minimum average block hour fleet utilization of 7 and 8 .................... 180
12.1 Estimation results .................................................................................. 191
12.2 Comparing solutions while varying the load factor multiplier ............. 201
12.3 Mean Absolute Percentage Errors ....................................................... 201
12.4 Comparing solutions for different scenarios for OD pairs La Coruña-Madrid and Madrid-La Coruña ..................................................... 213
Part I

INTRODUCTION
Chapter 1

OPERATIONS RESEARCH IN TRANSPORT ENGINEERING

This chapter gives a brief overview about the role that operations research play in transport planning.

1.1 Introduction

Transport is the foundation of any economy as it constitutes the heart of the supply chain. Without good transport networks, a proper functioning of the internal market is not possible. Transport infrastructure investments boost economic growth; create wealth; enhance trade, geographical accessibility and the mobility of people. They are a highly effective engine of job creation.

Transport is also a key ingredient for a high quality of life, making places accessible and bringing people together. Besides its role as a facilitator, the transport industry in itself represents an important part of the economy: in the European Union (EU) it directly employs around 10 million people and accounts for about 5% of Gross Domestic Product (GDP); many European companies are world leaders in infrastructure, logistics, traffic management systems and manufacturing of transport equipment.

Market integration, economic growth and transport activity are strongly related. In the EU, efficient transport connections have facilitated the creation and deepening of the internal market. Each of the EU enlargements was accompanied by a strong growth of transport activity.

The future prosperity of our world will depend on the ability of all of its regions to remain fully and competitively integrated in the world economy. Efficient transport is
vital in making this happen.

When operations research emerged as a structured field during World War II, some of the first problems investigated arose from the need to optimize military logistics and transport activities. After the war ended, the scope of operations research applications broadened but transport problems always occupied a central place (Barnhart and Laporte [10]). It is now widely recognized that some of the most successful applications of operations research are encountered in transport, most significantly in the airline industry where it underlies almost every aspect of strategic, tactical and operational planning. This success story may be explained by a number of factors, the first being the economic importance of transport. Also, the complexity and large scale of transport problems call for powerful analytical techniques, and the high volumes involved imply that substantial savings can often be achieved through the use of optimization. Furthermore, transport problems are highly structured, making them amenable to the use of efficient solution methods based on network optimization techniques and mathematical programming.

1.2 Planning in Transport

From a very abstract point of view, transport is caused by demands to move something between locations, say, from an origin to a destination. The demands fall into different categories: transport of people or goods, slow or fast, regional or inter-regional, periodically or spontaneous, etc. Transport systems are developed to satisfy these demands. It would be perfect to know all such demands for a foreseeable future. Then one could develop and plan transport systems that satisfy these requirements at lowest possible total cost. However, that is never possible.

It is easier to look at markets where transport systems have or are being developed in order to satisfy certain categories of demand. Each supplier of transport looks individually at 'its' market segment (e.g., airlines at long range traffic) and forecasts the relevant demand by estimating so-called origin-destination matrices which describe the amount of demand between locations (within certain demand-specific time intervals). Such forecasts are difficult to make on a long-term basis since they have to take local and global trends into account. Nevertheless, origin-destination matrices are estimated wherever quantitative methods have entered the decision process. Mathematics can help to some extent to improve the forecasting process and correctly interpret the estimated and computed data (see Codina and Barceló [50], Codina et al. [51] and García-Ródenas and Marín [67] for examples in adjusting origin-destination matrices).
Markets work by the interplay between supply and demand and there is, in particular, competition between technical systems. But the transport markets are not free since the state (or supranational units like the European Union) interferes and regulates in many ways: the state may support a new transport system within a certain market, e.g., to promote a new technology and create jobs. The state may also create almost monopolistic markets, often operated by a state-owned agency, as is frequently done for national railway systems or public transport within regions. This economic market view is a reasonable approach to make the acting forces - supply, demand, the state - visible. But nontrivial mathematics cannot contribute much to the solution of the non-quantifiable political, social, and economic decisions that come up at this high level.

Planning the design and operating existing transport systems, such as airlines, railroads or public transport systems, is the area where operations research enters the picture substantially. Let us consider one example, the planning process for the public transport system of a city. Three phases must be distinguished: a strategic, a tactical and an operational planning phase. Strategic planning has a planning horizon of several years or even decades, the main focus is on capacity planning; based on origin-destination data on the expected traffic demand between a number of representative points in a city, strategic decisions determine the amount of transport that a city is willing or capable to offer to its residents: the construction of new subway lines, the placement of depots, and the procurement of new vehicles are issues of strategic planning. Tactical planning allocates the available capacity with a planning horizon of 2 months up to a year: we may cite line planning and frequency planning among the most important problems in this phase. Finally, operational planning sets up the fully detailed plans, the planning horizon varies from 3 days to 2 months; mainly, it concerns adjusting the tactical plans to the forthcoming weeks: for example, timetable planning and vehicle scheduling give rise to a set of tasks to operate the individual vehicles that are next scheduled into duties in the crew scheduling step; then, the crews are assigned to pairings (i.e., sequences of duties) in a subsequent crew rostering phase. The problems arising in planning almost always come along with massive data and give rise to very challenging optimization tasks.

In the past, many areas of transport have been protected by monopolistic structures (typical examples are national railways or local public transport systems). Incentives for optimization were little. Consequently there was only low interest in the development of mathematical tools to aid in the decision process. But this situation has changed. In Europe, e.g., the deregulation of the transport sector by the Maastricht II treaty has put the transport companies under serious competition. Similar developments take place
elsewhere. A second point is of mathematical nature. Most transport problems arising in practice are really large. Hundreds of thousands of constraints and millions of variables are not uncommon. Problems of such dimensions were simply out of reach until very recently. Mathematical advances and advances in computing machinery, however, have changed the picture. Now it is possible to attack transport problems of sizes that were beyond imagination just a decade ago.

We provide here a summary of examples of the successful applications of operations research to transport. We mention a few examples where the operational planning was greatly improved by the use of mathematical optimization techniques (Borndörfer et al. [27]):

- **Airline Industry.** This is probably the most competitive sector in transport due to early deregulation. Operations research and optimization have a long history in this area; mathematical decision support techniques have been continuously employed and extended over the last 40 years. Leading companies use optimization techniques for daily, weekly, and monthly planning such as fleet assignment, crew scheduling, and crew rostering. Many airlines have created divisions or subsidiary companies to provide the necessary mathematical knowledge, consulting capacity, software tools, and computing machinery.

- **Rail Transport.** The use of mathematical methodology in this sector is not as widespread as in air traffic (the deregulation in this sector has not been so extended as in the airline industry); however, many railroad companies have recently observed the potential of mathematical decision support. From the operations research side, the planning process has been analyzed; models for several sub-problems ranging from line planning, timetabling, rolling stock and crew scheduling exist.

- **Public Transport.** For several decades now, operations research has been successful for solving a wide variety of optimization problems in public transit. Several commercial software systems based on operations research techniques have been designed and used by the transit agencies to help them plan and run their operations. The main goal of most transit agencies is to offer to the population a service of good quality that allows passengers to travel easily at a low fare. The agencies thus have a social mission which aims at reducing pollution and traffic congestion, as well as increasing the mobility of the population. The global problem faced by the agencies consists of determining how to offer a good-quality service to the passengers while maintaining reasonable asset and operating costs.
• Vehicle Routing. This is a classical area for optimization. But vehicle routing problems rarely come up in this 'pure' version in practice. Many legal and technical side constraints lead to a great variety of similar looking, but, in fact, quite different routing problems that have to be attacked with tailor-made solution methods. We should mention here also some closely related problems such as the sizing of vehicle fleets and the location of vehicle service facilities or distribution centers.

• Traffic Control. An important problem in road traffic is to manage the flow of vehicles such that congestion is small and vehicles are not forced to make big detours. Several ways to model the flow on highways and within cities have been proposed. These are based on cellular automata or analogies to fluid dynamics.

1.3 Recoverability and Robustness

The elaboration of a schedule starts long before the day of operations. Schedules are carried out in an uncertain environment (weather conditions, crew illness, technical failures, strikes, etc.), and often have to be adapted on the day of operations; events enforcing the schedule to be adapted are called disruptions. The direct consequences of disruptions are delays and cancellations, implying many schedule adjustments and huge costs. The consequences of a disruption are therefore amplified by both congestion and the sensitivity of optimized schedules: a local disruption propagates through the whole network, generating huge delays and, in consequence, huge recovery costs.

The recovery problem focuses on retrieving the original schedule from a disruption as soon as possible after it occurs while minimizing the recovery costs. This limits the consequences of the disruption. Such problems are, of course, particularly challenging, involving all resources, and requiring a global view of the system. Moreover, recovery decisions often need to be made in a matter of minutes. Hence, recovery methods are reactive processes based on a wait-and-see strategy for problems with varying data and are commonly called on-line algorithms. Such algorithms determine the response to data changes, which can either be deterministic or random. The performance of these algorithms is difficult to measure, as it depends on the way the data are revealed.

Consequently, it is crucial to develop schedules that are more robust, i.e. less subject to disruptions and also more recoverable, which means that, when a disrupted schedule has to be recovered, the induced recovery costs are minimized.

We can divide works on optimization for scheduling into distinct categories according to their objective: deterministic, stochastic, robust and recoverable robust methods.
- Deterministic methods are seeking an optimal solution according to a deterministic objective: possible uncertainty of the data is neglected. Unfortunately, as shown by numerous studies (e.g., Birge and Louveaux [24], Herroelen and Leus [74], Sahinidis [108]), deterministic optima are sensitive to small perturbations, which may make the solution inefficient or even unfeasible.

- Stochastic methods seek the solution performing best in average. There is usually an uncertainty set which describes the possible data realizations. One such realization is called a scenario. The uncertainty set is typically described by a probabilistic measure and the objective is to minimize the expected cost on all scenarios.

- Robust methods are more conservative than stochastic optimization, as they focus on the worst-case rather than the average case: the optimal solution is the one performing best in the worst possible scenario. Such methods are known as robust optimization, although the word stability may also be used. The pioneer of robust optimization was Soyster [114]; more recent works are provided by Bertsimas and Sim [23], Ben-Tal and Nemirovski [20], Ben-Tal and Nemirovski [21] and Ben-Tal and Nemirovski [22].

- Recoverable robust methods focus on producing solutions that are able to be recovered from a set of disrupted scenarios, using a restricted number of recovery algorithms in a limited effort (Liebchen et al. [89]). Consequently, we need to define a recovery algorithm in order to recover from each different disruption or scenario. The obtained solution is more flexible and can be adapted if necessary.
Chapter 2

THESIS OUTLINE

This dissertation is divided in four different parts. The first part presents a brief overview about the role that operations research play in transport planning; then, the thesis outline and contributions are presented. The second part of this dissertation presents the railway applications we have investigated. The third part the airline applications. Finally, the fourth part enumerates the conclusions of the research presented in this dissertation and the future research.

In the following sections, we briefly outline the contents of the chapters in this dissertation.

2.1 Chapter 3: Thesis Contributions

This chapter briefly describes the main contributions to the literature of the research presented in this dissertation.

2.2 Chapter 4: Introduction

A brief overview of the railway industry and planning process is provided in this chapter.

2.3 Chapter 5: Robust Rolling Stock in Rapid Transit Networks

In this chapter, we study the Rolling Stock (RS) problem at the operative level in rapid transit networks; such networks operate in metropolitan areas, and feature frequent train
services and heavy passenger loads. We focus on the detailed planning to be operated in a daily basis. The aim is to build a daily schedule that minimizes operating costs while accounting for passenger demand. For that purpose, empty services and shunting must be included in the problem. Moreover, robustness is introduced in the problem by ameliorating difficult shunting operations. The problem is modeled as a large-scale, network-based, mixed-integer programming (MIP) problem.

All the analysis is based on extensive amounts of confidential data such as railway schedules and passenger flows given by RENFE, the main rail operator in Spain. The results obtained are satisfactory: operating costs are lowered while a high level of service quality for passengers is maintained and robust plans for network operation are provided. Moreover, the time needed to obtain these plans is reduced from the current system of manual planning under great time pressure.

For a deeper insight of the problem presented in this chapter see Cadarso and Marín [34].

2.4 Chapter 6: Robust Routing of Rapid Transit Rolling Stock

The train routing problem determines the sequence for specific train units. In other words, once we know the train unit type assigned to each operation (namely the RS assignment), we must know which operation precedes and succeeds it.

In this chapter, we develop a robust train routing model to determine train units’ sequences in a rapid transit network that attempts to minimize the delay propagation in each sequence as well as the crew requirements at depot stations. Here, robustness means that conflicting material connections are spread out in time as much as possible.

We report our computational tests on realistic problem instances of RENFE. They are achieved for short times and, based on a previous efficient RS assignment, show that a more robust and efficient solution than the current one can be obtained.

For a deeper insight of the problem presented in this chapter see Cadarso and Marín [33].
2.5 Chapter 7: Integration of Timetable Planning and Rolling Stock in Rapid Transit Networks

The railway planning process has been historically sequentially solved. This approach allows the planners to divide the whole planning process into smaller problems. However, this disintegrated planning does not provide the global optimum to the planning problem. This issue may lead to the system to operate in an inefficient way. The integration increases the flexibility and the robustness of the railway system and also makes easier to introduce robustness to the system.

In this chapter we match demand requirements updating frequencies, planning trains’ departure times and assigning RS train units for a given planning period in a rapid transit network. For RENFE, the planning period is of one day, which is repeated during the working days of a week, so a periodic planning is needed where the data and the decisions must be considered in the context of a space-time network.

Computational experiments based on RENFE problem instances show how the RS assignment efficiency depends on the timetable. Better plans can be obtained when the timetable planning and RS assignment problems are integrated.

For a deeper insight of the problem presented in this chapter see Cadarso and Marín [36].

2.6 Chapter 8: Improving Robustness of Rolling Stock Circulations in Rapid Transit Networks

The RS circulation depends on two different problems: the RS assignment and the train routing problems, which up to now have been solved sequentially.

In this chapter, we propose a new approach to obtain robust RS circulations in a rapid transit network: it consists of solving the RS assignment and the routing problems in an integrated way. This integrated approach provides a huge model. Therefore, we propose a heuristic based on Benders decomposition.

Computational experiments show how the current solution operated by RENFE can be improved: more robust and efficient solutions are obtained.
2.7 Chapter 9: Recovery of Disruptions in Rapid Transit Networks

During the daily operations of a dense railway network, incidents may cause the railway traffic to deviate from the planned operations. These incidents may make it impossible to operate the schedule as it was originally planned. In such a situation the operator needs to adjust the timetable and the RS assignment for the time interval of the incident, and to carry out further recovery steps in order to get back to the original schedules.

We study the disruption management problem of rapid transit rail networks. Besides optimizing the timetable and the RS schedules, we explicitly deal with the effects of the disruption on the passenger demand. We use a multinomial logit model to anticipate the disrupted passenger demand. Then, we solve an integrated Mixed Integer Linear Programming model for the timetabling and RS scheduling problem. We investigate an iterative framework around our integrated optimization model where the demand of the next iteration is computed from the optimized timetable of the current iteration.

We report our computational tests on realistic problem instances of RENFE. The proposed approach is able find solutions with a very good balance between various managerial goals within a few minutes.

For a deeper insight of the problem presented in this chapter see Cadarso et al. [38].

2.8 Chapter 10: Introduction

A brief overview of the airline industry and planning process is provided in this chapter.

2.9 Chapter 11: Robust Passenger Oriented Timetable and Fleet Assignment Integration in Airline Planning

The motivation behind this chapter stems from the fact that, up to now, the planning phases have been solved in a sequential manner providing suboptimal and even infeasible solutions: scheduling connecting itineraries is a difficult task related to flight times, flight capacity and passengers.

Therefore, we look at the airline-scheduling problem and develops an integrated approach that optimizes schedule design, fleet assignment and passenger use so as to reduce
costs and create fewer incompatibilities between decisions. Robust itineraries are created to ameliorate misconnected passengers.

The analytical work is supported with a case study involving the Spanish airline, IBERIA. Our approach shows that the number of misconnected passengers can be reduced when robust planning is applied.

For a deeper insight of the problem presented in this chapter see Cadarso and Marín [37].

2.10 Chapter 12: Integrated Airline Scheduling: Considering Competition Effects and the Entry of the High Speed Rail

Airlines and High-speed rail are increasingly competing for passengers. Competition between them affects the number of captured passengers and, therefore, revenues. We consider multi-modal competition including airline (legacy and low cost) and high-speed rail, and develop a new approach that estimates the demand associated with a given schedule using a nested logit model, and generates airline schedules using an integrated optimization model that captures the impacts of decisions on passenger demand. Our experimental results show that we are able to replicate IBERIA’s (the major Spanish airline) current decisions with a reasonable level of accuracy, thus validating our modeling approach. We evaluate multiple scenarios involving entry of High-speed rail in some markets, and we account for the possibility of demand stimulation as a result of the new services.

2.11 Chapter 13: Thesis Conclusions

This chapter presents the main conclusions of this dissertation for each chapter in a separate way.

2.12 Chapter 14: Future Research

This chapter presents future research we are going to embark in the near future.
Chapter 3

THESIS CONTRIBUTIONS

In the following, we describe the major contributions of this thesis to the existing literature. We detail the contributions of each chapter in a separate way.

3.1 Chapter 5: Robust Rolling Stock in Rapid Transit Networks

In this chapter, we present a mixed integer and linear programming model for the rolling stock assignment problem that includes shunting maneuvers in rapid transit networks. The major contributions include:

1. Development of a multicommodity flow model in which commodity (namely, train units) flows are studied. Once these flows are known, train sequences will be determined in the routing problem.

2. Introduction of different quality of service levels: capacity for standing passengers is not fixed and we account for different configurations.

3. Introduction of shunting maneuvers; they are necessary to operate rapid transit networks where resources are limited and frequency values are high.

4. Development of robust plans: they ameliorate the negative effects of disruptions. We introduce robustness in the rolling stock assignment problem selectively avoiding operations that may malfunction or that may be difficult to be performed (namely, composition changes and empty movements).
5. Evaluation of the proposed mathematical model in the Madrid RENFE suburban rail network. Satisfactory results are obtained: operating costs are lowered, high level of service quality for passengers is maintained and robust plans are provided. The time needed to obtain these plans is reduced from the current system of manual planning.

3.2 Chapter 6: Robust Routing of Rapid Transit Rolling Stock

We present a model to study the routing problem in a dense metropolitan network, where no literature addressing rapid transit networks has been found. The major contributions include:

1. Development of a sequencing model in which train units’ sequences are studied.
2. Robustness is introduced spreading conflicting train connections out in time. Expected train service delays and the consequent number of passengers at stations are minimized.
3. Efficient matching of operations in order to minimize crew requirement. Extra human resources are necessary to perform some operations in depot stations. We introduce a time penalization to avoid these connections, reducing crew requirements and introducing robustness.
4. Computational tests on realistic instances of the Spanish rail operator RENFE are performed. Expected delay is sensitively reduced: expensive swapping operations are reduced. Crew resources are also minimized.

3.3 Chapter 7: Integration of Timetable Planning and Rolling Stock in Rapid Transit Networks

In this chapter, we present an integrated model to study the timetable development and the rolling stock assignment. The major contributions include:

1. Integration of the timetable development and the rolling stock assignment planning phases in one single model. Moreover, frequency values are not known: minimum and maximum frequency values are imposed and the model calculates them.
2. Introduction of detailed shunting maneuvers.

3. Introduction of robustness by penalizing dangerous shunting operations and ensuring swapping operations in strategic depot stations.

4. Evaluation of the proposed mathematical model in the Madrid RENFE suburban rail network. The efficiency of the rolling stock depends on the timetable: comparison of the presented approach with the rolling stock assignment with fixed timetable demonstrates how better solutions can be obtained improving the overall solution efficiency and getting a greater robustness degree.

3.4 Chapter 8: Improving Robustness of Rolling Stock Circulations in Rapid Transit Networks

In this chapter, we present a model to study the rolling stock assignment and routing problems in an integrated way. The major contributions include:

1. Integration of the rolling stock assignment and train routing planning phases in one single model.

2. Introduction of robustness by minimizing expected delays and swapping operations (swapping operations are performed in order to avoid the downstream effects of a delay).

3. Development of a Benders based heuristic in order to solve the proposed mathematical model.

4. Evaluation of the presented mathematical model and solving method in the Madrid RENFE suburban rail network. Robust and smooth solutions are obtained: delays and dangerous shunting operations are ameliorated.

3.5 Chapter 9: Recovery of Disruptions in Rapid Transit Networks

In this chapter we present a new approach to deal with large-scale disruptions in rapid transit networks. The major contributions include:
1. Application of a multinomial logit model to compute the anticipated passenger.

2. Development of an optimization model to be applied in case of disruption that simultaneously deals with timetabling and rolling stock scheduling decisions subject to the anticipated demand.

3. Investigation of an iterative framework around the integrated optimization model where the demand of the next iteration is computed from the optimized timetable of the current iteration.

4. Evaluation of the presented approach on realistic instances of the Spanish rail operator RENFE: the method is able to find solutions with a very good balance between the different managerial goals. Computational times amount to a few minutes which is sufficiently close to the needs of real-time decision making.

3.6 Chapter 11: Robust Passenger Oriented Timetable and Fleet Assignment Integration in Airline Planning

We present a new integrated approach to solve two key planning phases in the airline industry: the timetable planning and fleet assignment. These are tactical problems where a planning horizon of several months or one year is available for planners. The major contributions include:

1. Development of an integrated optimization and itinerary based schedule design model that includes timetable development and fleet assignment.

2. Passenger demand is realized through itineraries in the airline network. Introduction of robustness in the model through passengers’ itineraries, avoiding misconnected passengers due to lack of time to perform connections. Passenger recapture is included.

3. Fleet utilization aspects are included in the model: a minimum average block hour utilization is mandated for every fleet type. Different scenarios are studied by varying it.

4. As a proof of the model computational experiments in a real network from the Spanish airline IBERIA are shown. The robust approach is compared with a non-robust
one: more robust solutions can be achieved by reducing the number of expected
misconnected passengers.

3.7 Chapter 12: Integrated Airline Scheduling: Considering Competition Effects and the Entry of the High Speed Rail

In this chapter, we present a mixed integer and non-linear programming model for the schedule design and fleet assignment problem that includes a passenger choice model to capture multi-modal competition between high-speed rail, low-cost airlines, and legacy airlines. The major contributions include:

1. Development of a tactical competition model for an airline-considering multi-modal competition between air and high-speed rail, and airline competition between legacy and low cost carriers-using a nested logit model of demand behavior. We calibrate the model using real data.

2. Development of an integrated optimization schedule design model that includes frequency planning, approximate timetable development, fleet assignment and passenger demand choice.

3. The model is tested using realistic problem instances obtained from the network of IBERIA, and including other air and rail transportation options in Spain, and perform sensitivity analyses on various model parameters.

4. Evaluation of multiple scenarios involving entry of High Speed Rail in some markets: demand stimulation as a result of the new services is studied.
Part II

RAILWAY APPLICATIONS
Chapter 4

INTRODUCTION

The objective of this chapter is to provide an overview of the railway industry and planning process.

4.1 The Railway Industry

Railway systems are highly complex systems. Planning and operational processes related to railway systems are fields that are rich in interesting combinatorial optimization problems. Well-known examples of these problems are planning problems such as network design, line planning, timetable planning, rolling stock assignment, routing, and crew planning (Caprara et al. [45]).

However, in the railway industry it was recognized only recently that such problems can be analyzed and solved through the application of mathematical models and optimization techniques, and that this kind of innovation may lead to an improvement in the performance of the railway system as a whole, as well as to a reduction in the time required for solving these problems. The latter aspect is important, because it increases the flexibility of the railway system: the system can adapt in a faster way to changes in the environment.

Railway transportation can be split into passenger transportation and cargo transportation. In this work, we mainly focus on passenger transportation in rapid transit networks. Whereas in the recent past many European railway companies were state-owned, many of them are currently operating (partly) on a commercial basis, due to the new regulations of the European Commission, which specify that the management of the infrastructure should be the responsibility of the governments, but operating trains should be carried out by independent train operators on a commercial basis. This introduces the
separate organizations of the infrastructure manager, who is responsible for train planning and real-time traffic control, and the train operators, who provide their preferred timetable and transport services.

The recognition that combinatorial optimization problems arising in railway applications can be solved through the application of mathematical models and of the corresponding solution techniques was caused by several factors. Indeed, the ability to undertake infrastructure planning in a very timely, smooth and efficient way is becoming one of the most important tasks of the infrastructure manager, who at the same time has to optimize the use of the infrastructure and to provide line allocation as well as time slot allocation. Moreover, the train operators increased the demand for improved performance and for higher speed and flexibility in the planning and in the operations. This also stimulated the search for innovative decision support tools, which were recognized much earlier already in the airline industry.

Railway companies’ planning problems have been dealt manually till recently; many are still handled without automation and optimization. Many of the problems are of a combinatorial character and suitable for operations research methods. In the last decade, more and more computer-aided tools turned out to improve the railway planning process significantly. Besides intensive research, the exponential increase of computational power contributes a lot to these successful applications.

For many years, railway companies did not have to face much competition in public passenger and freight transportation. In the past decades, this changed drastically. The railways lost a large part of their market share to automobiles. Recently, air traffic took over many middle- and long-distance train travelers. In addition, a directive of the European Union required opening the national railway market in the 90’s. Till then, most state-owned railway companies in Europe had been the only ones to provide railway services in their countries; after liberalizing the market, they had to compete for the customers. These developments urge the railway companies to attract more customers by raising their service level and to cut their costs by working more efficiently. Improving their planning process contributes to reach both of these goals.

Planning the railways for years, months, weeks or days ahead leads to substantially different problems; in this regard railway planning problems can be strategic, tactical, operational and short-term. Strategic planning has a planning horizon of several years or even decades, the main focus is on capacity planning. Tactical planning allocates the available capacity with a planning horizon of 2 months up to a year. Operational planning sets up the fully detailed plans, the planning horizon varies from 3 days to 2
months. Mainly, it concerns adjusting the tactical plans to the forthcoming weeks. Short-term planning comprises problems that arise when the actual train operations take place or just before that (i.e., the recoverability of the system when a disruption occurs). It covers planning steps with a planning horizon of at most 3 days.

4.2 The Railway Planning Process

In this section the main railway planning phases are introduced. They are all based in the network design developed by the infrastructure manager (see Gendron et al. [69], Marín and García-Ródenas [92] and Marín et al. [93] for an examples in network design problems).

The most important planning decisions faced by railway managers can be summarized as follows:

- Fleet planning
- Line planning
- Timetable development
- Rolling stock assignment
- Rolling stock routing
- Crew assignment

More tactical decisions related to marketing and distribution are also required, involving pricing and revenue management.

In the following subsections each of the mentioned planning problems above are shortly introduced.

4.2.1 Fleet Planning

Strategic rolling stock planning aims at determining the number of train units (self-propelled units with driver seat at both ends; they may be of different types, i.e. passenger capacity) needed to cover passenger demand in the forthcoming years (Maróti [94]). It takes years until newly ordered units are actually delivered. The expected service time of the units amounts to decades. Strategic decisions on rolling stock concern large amounts of money and they determine the train traffic and the service quality for years.
The basic decisions are purchasing or leasing new rolling stock and refurbishing existing units. Redundant units may be sold or disassembled. Another problem is to set the strategy for regular preventive maintenance. More frequent maintenance probably results in more reliable rolling stock, decreasing the disturbances caused by defect units. On the other hand, it also leads to higher maintenance costs and it requires a higher number of units in use.

4.2.2 Line Planning

A line is defined by trains with the same route and stop stations. The line design considers the demand satisfaction and some capacity constraints. There are two main conflicting objectives to be pursued when planning a line system, namely (i) optimizing passenger service (Bussieck et al. [30], Scholl [110], Schöbel and Scholl [109]), and (ii) minimizing operational costs of the railway system (Claessens et al. [47], Cordeau et al. [52]).

The improvement of the passenger service may be defined from different points of view: minimizing transfers, minimizing total travel time or maximizing comfort. Maximizing the number of direct trips usually results in long lines, however, long lines may transfer delays more easily and provide an inefficient allocation of rolling stock, because it is usually allocated according to the peak demand along the line. Therefore, in a robust system the lines are relatively short, which may force passengers to transfer from one train to another too often.

When designing a line system, one has several options for providing sufficient capacity to transport all passengers: lines can be operated with a high frequency and with trains with a relatively low capacity, or with a low frequency with trains with a relatively high capacity.

Finally, it should be noted that in practice there is a bi-directional relation between the travel demand and the operated line system. On the one hand, the operated line system should be such that the full travel demand can be accommodated. On the other hand, a line system that provides a high service towards the passengers may also attract additional passengers, thereby increasing the original travel demand. Thus the travel demand also depends on the operated line system.

4.2.3 Timetable Planning

The general aim of the train timetabling problem is to provide a timetable for a number of trains on a certain part of the railway network. One may distinguish between cyclic
timetables (Nachtingall [98], Nachtingall and Voget [99], Kroon and Peeters [83]) and non-cyclic timetables (Carraresi et al. [46], Brännlund et al. [25], Caprara et al. [43], Caprara et al. [44]).

An advantage of a cyclic railway system is the fact that such a system’s timetable is easy to remember for the passengers. For example, at a certain station, the trains heading for a certain direction always leave at :12 and :42. On the other hand, a drawback is that such a system is expensive to operate. Even in the off-peak periods with low travel demand, more or less the same timetable is operated as during the peak hours. The only way to differentiate the system’s capacity between the peak hours and the off-peak hours is to modify the lengths of the trains. The latter impacts both the variable rolling stock costs and the variable crew costs.

The non-cyclic timetable is especially relevant on heavy-traffic, long-distance corridors, where the capacity of the infrastructure is limited due to greater traffic densities, and competitive pressure among the train operators is expected to increase in the near future. This allows the infrastructure manager to allocate ‘optimally’ the train paths requested by all train operators and proceed with the overall timetable design process, possibly with final local refinements and minor adjustments, as in the tradition of railway planners. In brief, this allows each train operator to submit requests for paths on the given railway line, and allows the infrastructure manager to collect all the requests, run the optimization algorithm to allocate (if possible) all of them at maximum profit, and eventually respond to the train operators with the proposed plan of the time slot allocation and the relative ‘access fees’.

4.2.4 Rolling Stock Assignment

Rolling stock planning determines how many train units are needed and how to use them for trains. The major objectives in rolling stock planning are service quality and operational costs (Maróti [94]). A good service quality means that trains have enough capacity to cover the passenger demand. A higher service quality encourages more travelers to use the train instead of their cars. When running the trains, railway operators have rolling stock related expenses such as electricity or fuel consumption and maintenance costs; efficient schedules minimise these expenses.
4.2.5 Rolling Stock Routing

The solution obtained from the rolling stock assignment problem identifies the flow of different train units through the rapid transit network. However, it does not identify which specific train unit is assigned to each operation (service to be performed). Rolling stock routing is the process of assigning each individual train unit, referred to as an identification number, to every operation. Given an assignment of train unit types to operations, we must determine a sequence of operations to be rolled by an individual train unit such that every operation is included in exactly one sequence, and there are always enough resources available for every operation.

This approach is widely extended in the airline industry. Airlines usually solve the fleet assignment problem (equivalent to the rolling stock problem). Then, the aircraft routing problem is solved. Lan et al. [85] propose an aircraft-centric model minimizing propagated delay by routing aircraft.

4.2.6 Crew Assignment

The crew planning problem is the problem to be faced by train operators which is concerned with building the work schedules of crews needed to cover a planned schedule. We are given a planned schedule for the train services (i.e., both the actual journeys with passengers, and the empty trains or equipment between different stations) to be performed every day of a certain planning horizon. Each service has to be performed by one crew. In fact, each crew performs a roster, defined as a sequence of services whose operational cost and feasibility depend on several rules laid down by union contracts and company regulations. The problem consists of finding a set of rosters covering every service of the given planning horizon, so as to satisfy all the operational constraints with minimum cost.

Usually, the crew planning problem is approached in two phases, according to the following scheme:

1. Crew Scheduling: the short-term schedule of the crews is considered, and a convenient set of duties (pairings) covering all the services is constructed. Each duty represents a sequence of services to be covered by a single crew within a given planning horizon.

2. Crew Rostering: the duties selected in the crew scheduling phase are sequenced to obtain the final rosters.
Abbink et al. [2] deal with large-scale crew scheduling problems. They discuss several methods to partition huge instances into several smaller ones. Hartog et al. [72] describe a method for solving the cyclic crew rostering problem. This is the problem of cyclically ordering a set of duties for a number of crew members, such that several complex constraints are satisfied and such that the quality of the obtained roster is as high as possible.

4.3 Railway Planning Problems at RENFE

RENFE is the main Spanish train operator. It is a state-owned passenger and freight rail transport operator overseen by the Development Ministry. It provides passenger and freight transport services. It also provides service to passengers in the suburban areas. RENFE operates rapid transit networks in these areas, which are fully devoted to passenger transportation.

Some of the above presented problems will be addressed during this dissertation. All of them will focus on the rapid transit network in Madrid, Spain, which is fully described here. All of our computational experience is for realistic cases using data of 2008.

4.3.1 The Rapid Transit Network

Rapid transit networks are characterized by high frequencies and short distances. However, service times may be large even though distances are short. This is due to the large number of intermediate stops required to meet passenger demand. Figure 4.1 shows the rapid transit network in Madrid, Spain.

In this type of networks, we can distinguish two different principal types of stations. The first type is characterized by train services that only serve to passenger demand. The second type is called a depot station. In these stations, shunting operations can also be performed. That is, attached to the passenger station is a depot where trains are driven to be parked or shunted. Some of the depot stations are specially characterized because they share their capacity among different lines. Because of this capacity sharing, their capacity may change during different time periods, complicating shunting operations. This time-dependent capacity is determined in a previous planning phase jointly with the infrastructure manager.

The existing infrastructure linking different stations is represented by arcs. Between two stations, two different arcs exist, one for each direction of movement. Therefore, every arc is defined by its departure and arrival station and by its length (e.g., in kilometers).
The planning time is discretized into time periods. Due to the high train frequencies, the duration of one time period is set to one minute. The existing physical network is replicated once for each time period existing in the planning period (e.g., 20 hours).

4.3.2 Timetable

Train services are grouped in lines. A line is characterized by its terminal stations, by a path through the infrastructure between the terminals, and by a set of stations along the path. Train services run up and down between the terminals and call at the specified stations underway. Note that, under exceptional circumstances, a service may operate on a subpath only; this happens early in the morning, late in the evening, and also during a disruption.

The timetable departure times and frequencies are fixed and publicly available. The passengers know when the trains depart and plan their traveling accordingly. Departure times are very inflexible because the time slots are negotiated with a third party (the infrastructure manager) since the network infrastructure is shared among different lines. However, for disrupted situations there is some freedom to schedule services with different timetables.
A planned train service is a passenger train traveling from a depot station to another depot station stopping at a number of intermediate stations. They are characterized by their departure depot station; their arrival depot station; every arc they travel on; and their departure time. The distance rolled by a train service is the sum of the lengths of the arcs used by the train service.

Rapid transit networks are characterized by high frequencies and a lack of capacity in depot stations. These facts make it difficult to operate the network without empty movements. These are defined by an origin, a destination and a departure time. Empty movements can help satisfy both capacity and rolling stock availability in depot stations.

4.3.3 Rolling Stock and Shunting

There are self-propelled train units of different types; they all have a driver seat at both ends. Units of the same type can be attached to each other to form trains compositions.

Shunting operations such as rotations and composition changes complicate rapid transit networks because the performance time is on the order of the service frequency time.

Every departing train service must have time to perform a rotation. That is, the composition assigned to a train service must be available at the depot station a certain number of periods before the departure time. This is due to the fact that train units need some time to be prepared and functional for departure.

Train units of the same type can be aggregated to form longer compositions, and compositions can be disaggregated into individual train units. Although composition changes enable the network operator to use smaller fleet sizes, it is always a complicating operation, due to the necessity of human resources and the possibility of failure in the mechanical system governing the process.

4.3.4 Passenger Demand

The passenger demand in this dissertation will be treated as a passenger flow through each arc belonging to each train service. This passenger flow is obtained from historical data under normal conditions (i.e., assuming that the train services matched the designed timetable). In this case, the passenger flow is known and is used to fix the required capacity for each arc in each service, assuming that the passenger flow carries complete information about the train service timetable and the capacity that will be offered (i.e., assuming equilibrium has been reached).

Under the above hypothesis, we will treat the passengers from a centralized point of
view (i.e., only the operator criteria are optimized). However, since the proposed problem relates to a suburban rapid transit network, it is obvious that every passenger will have the option to choose any other available company or transportation mode. Thus, the operator has to factor in passenger behavior to avoid losing passengers to other transportation companies.

The first and best choice would be to include the actual passenger behavior. Although this approach would give almost the true optimum, it is very complicated to include it in the model in a realistic way due to the large number of possibilities involved in rapid transit networks and their competitive operators. In a simplified approach, passenger behavior can be summarized as follows: if the passenger maintains his/her satisfaction with the transportation mode, he/she will remain in the system. As long as the system operator maintains certain standards within the transportation system, we assume that the passenger flow is known (i.e., the equilibrium mentioned above is achieved).

Under the assumption that public timetables are met, transportation standards might be described by the capacity offered in each train service. The operator will always try to offer comfortable capacity in each train service. It is obvious that when more capacity is offered, more passengers will travel comfortably. However, offering more capacity increases operating costs dramatically. Therefore, the composition assigned to each train service will be a tradeoff between the operating costs and the behavior of passengers (represented by their comfort level).

For each train unit, the passenger capacity is known. There is a fixed seating passenger capacity and a variable standing passenger capacity. Multiple possibilities arise when considering the standing passenger capacity. The aim is to obtain adequate passenger capacity for every train service. This may be obtained with different configurations for standing passengers.

4.3.5 Part II Outline

In the following chapters some railway optimization problems will be addressed. All of them have been developed and tested in the rail rapid transit network presented in 4.3. The presented problems are:

- the robust rolling stock assignment,
- the robust routing of the rolling stock,
- the integration of the timetable planning and rolling stock assignment,
• the integration of the rolling stock assignment and routing and,

• the recovery of disruptions.
Chapter 5

ROBUST ROLLING STOCK IN RAPID TRANSIT NETWORKS

This chapter focuses on the railway rolling stock circulation problem in rapid transit networks. The main complicating issue is the fact that the available capacity at depot stations is very low, and both capacity and rolling stock are shared between different train lines. This forces the introduction of empty train movements and rotation maneuvers, to ensure sufficient station capacity and rolling stock availability.

However, these shunting operations may malfunction, causing localized incidents that could propagate throughout the entire network due to cascading effects. This type of operation will be penalized with the goal of selectively avoiding them and ameliorating their high malfunction probabilities.

We illustrate our model using computational experiments drawn from RENFE in Madrid, Spain. The results of the model, achieved in approximately one minute, have been received positively by RENFE planners.

5.1 Introduction

In a daily planning period, the data and the decisions must be considered in the context of a space-time network. A known demand and timetable are met by a given fleet. The Rolling Stock (RS) model makes decisions about the aggregation and disaggregation of the different RS in the depot stations; in this phase, shunting operations are studied in detail and generic plans (the result of tactical planning) may be adjusted to meet the specific demands of particular scenarios. The problem can be stated as follows in the context of metropolitan rapid transit networks: given the train services’ departure and
arrival times and the expected numbers of passengers at each arc and in each period, and considering shunting operations, find the optimal assignment of the RS to the train services.

Major complicating issues are the available shunting capacities at depot stations and RS sharing between lines. Shunting is related to the need for RS to be parked in shunting areas when the RS is not used for traffic during off-peak hours and for those maneuvers to match compositions during times between the beginning and end of the planning period. The shunting process is very complex for urban and suburban depots because depot stations are shared between RS moving on different train lines. This implies that depot station capacities may change in different time periods. This forces us to combine and split train units to form trains and to consider the logistics of empty train movements in order to meet depot station capacities. However, empty trains will also be moved to ensure RS availability because RS is a very limited resource during rush hours, when the passenger demand is very high.

As we have stated above, both RS resources and depot station capacities are limited. Thus, we introduce composition changes to improve the availability of RS. Another complicating issue is rotation times. Rotations are the maneuvers performed at depot stations to change the direction of motion of the RS. It is assumed that the service time is the actual service time plus the rotation time; this time is known in the literature as the availability time. However, in rapid transit networks, in which capacities are limited and frequencies are on the order of the rotation time, we need to account for the depot station capacity in each period to avoid exceeding capacity.

This chapter presents a specialized RS model (RSM) describing rapid transit networks. The RSM will consider the optimization of train services’ compositions, empty trains and the optimal management of train units in depot stations, all while considering the character and capacities of these types of RS and depots. The RSM is a first approach to the new subject of urban rapid transit networks, which, have been manually planned to date.

As stated above, empty movements and shunting operations are necessary in rapid transit networks. However, these operations may sometimes be difficult to execute and they can easily malfunction, causing localized incidents that could propagate through the entire network due to cascading effects. This is the case for composition changes. These operations will be penalized to selectively avoid them and their high malfunction probabilities. Alternatively, we can introduce robustness by avoiding empty train movements. During rush hours, the network is very congested and an empty movement using the same
infrastructure as commercial trains increases the probability of an incident in the network. Empty movements at rush hours will also be penalized.

This chapter is organized as follows. A literature review is given in Section 5.2. We describe the RS problem for rapid transit networks in Section 5.3. In Section 5.4, the mathematical formulation is presented in detail. Section 5.5 contains the computational results based on a realistic case provided by RENFE.

5.2 Literature Review

Schrijver [111] minimizes the number of train units that must be employed in a line to avoid standing passengers. Although a unique type of train unit is considered, two subtypes can be defined. There is only one constraint for composition changes: between two consecutive trips, a change is allowed if the required vehicles are in the right place at the right time. An integer programming model is considered by Alfieri et al. [3] to determine the RS circulation for multiple RS types on a single line and on a single day, taking into account the order of the units in the train compositions. They use the concept of a transition graph to deal with this aspect. This concept is based on the assumption that for each trip, the next trip is known a priori. The problem is an integer multi-commodity flow problem, in which a feasible path in the transition graph is to be found simultaneously for each train. The objective is to minimize the number of units or the carriage-kilometers such that the given passenger demand is satisfied. The approach is tested on real-life examples from NS, the main operator of passenger trains in the Netherlands. The model described by Alfieri et al. [3] was extended by Fioole et al. [63], to include combining and splitting trains, as happens at several locations in the Dutch timetable. Robustness is considered by counting the number of composition changes. Maróti [94] focuses on planning problems that arise at NS. He identifies tactical, operational and short-term rolling stock planning problems and develops operations research models for describing them. Then, he analyzes the considered models, investigates their computational complexity and proposes solution methods. The allocation of RS units to French TGV trains is studied by Ben-Khedher et al. [19]. The RS circulation must be adjusted to the latest demand known from the seat reservation system. The objective is to maximize the expected profit for the company. A locomotive and carriage assignment problem was presented by Cordeau et al. [53]. The authors formulate the problem as a large integer program and use Benders decomposition to solve it. In a subsequent paper by Cordeau et al. [54], their model was extended by considering various aspects such as maintenance of the RS. A RS circulation problem
related to the circulation of ICE train units in the German ICE network was described by Mellouli and Suhl [95]. In this case, the required capacities of the trains are known a priori. Carriages and locomotives first have to be combined into train units of certain pre-specified groups, and these train units then have to be routed through the network in an optimal way. The problem is modeled as an integer multi-commodity flow problem on a multiple-layered network. Marín and Cadarso [91] define a model to study suburban rapid transit RS with convoys formed by three cars of the same type. The trains may be composed of one or two convoys in a dense network to attend to asymmetric demand and scheduling.

An elaborate introduction to the shunting problem including a solution approach and computational results, can be found in Freling et al. [64] and Lentink [86]. In these papers, the matching and parking subproblems are solved separately. The problem of routing RS units between the platform and shunting areas of a station was studied by Van den Broek et al. [120].

Liebchen et al. [88] introduced the concept of the Price of Recoverability as a generic framework for modeling robustness issues in railway scheduling problems. However, the notion is theoretical in nature and is not straightforward to use in specific problems. Cacchiani et al. [39] and Cacchiani et al. [40] explore the application of the Price of Recoverability and Recoverable Robustness ideas to railway RS planning. They are particularly interested in practically computable recoverability measures, thereby evaluating the robustness of real-life RS schedules. Their focus lies in real-life resource scheduling problems that are formulated as mathematical programs. They evaluate the approach on real-life RS planning problems encountered by NS.

The RS problem is very similar to the fleet assignment problem in the airline industry. Different aspects of this problem have been deeply researched. Daskin and Panayotopoulos [55] present an integer program that assigns aircraft to routes (which they define as sequences of flight legs originating and terminating at the same airport) in single-hub networks. Lagrangian relaxation is used to find an upper bound, and heuristics are used to find specific solutions. Abara [1] presents a model that can be used in more general airline networks, but the model has some limitations due to the use of connection arcs as decision variables. The model size explodes unless limits are placed on the connection opportunities. Another limitation is that different flying times and turn times (minimum ground service times) are not allowed for different fleets. Hane et al. [70] present a multi-commodity flow model. They show a number of ways to reduce the problem size: variable aggregation, cost perturbations, dual simplex with steepest-edge pricing, and
intelligent branch and bound strategies. Fleet assignment models have been widely applied in practice and costs have been significantly reduced as we can see in Delta Airlines (Subramanian et al. [116]).

5.3 The Rolling Stock Problem in Rapid Transit Networks

The rolling stock problem in this chapter has been studied for the rapid transit network introduced in [4.3]. There, a physical network composed of stations, indexed by \( s \in S \), and arcs, indexed by \( a \in A \) is presented. The planning time is discretized into time periods, indexed by \( t \in T \), and the physical network is replicated in time obtaining a space-time network.

**Train Services.** Train services which are known operate within the space-time network. Each train service is represented by \( \ell \in L \). Train services are defined as commercial trains operating in the network to meet passenger demand. They are characterized by their departure depot station; their arrival depot station; every arc they travel on; and their departure time. The length of a train service \( \ell \) is the sum of the lengths of the arcs used by the train service. Consequently, an indicator \( \alpha_{\ell,s,t} \) containing the timetable is known (this timetable satisfies headway requirements). It will take the value -1 (1) if train service \( \ell \) departs (arrives) from (at) depot station \( s \) during time period \( t \).

**Rolling Stock.** Each train service \( \ell \) will have to be assigned a RS composition \( c \in C \) of a determined type \( m \in M \). A composition is defined by the number of train units composing it. In the rapid transit network of Figure [4.1] material is not mixed within the same composition, i.e., the same material is used for building every composition. However, compositions of different types may be assigned to different train services.

**Passenger Demand.** For each train unit of material type \( m \), the passenger capacity is known. There is a fixed seating passenger capacity and a variable standing passenger capacity. Multiple possibilities arise when considering the standing passenger capacity. The aim is to obtain adequate passenger capacity for every train service. This may be obtained with different configurations for standing passengers. We define comfortable capacity as the full seating capacity and fewer than 3 \( \text{pax/m}^2 \) standing. If this capacity is exceeded, the passengers above this level are deemed passengers in excess. If the density of standing passengers is between 3 and 4 \( \text{pax/m}^2 \), each passenger in excess represents a moderate penalty, because the operator would like to obtain 3.5 \( \text{pax/m}^2 \). If the density of standing passengers exceeds 4 \( \text{pax/m}^2 \), passengers in excess are highly penalized because
this situation is deemed very uncomfortable.

The latter can be formulated as a piecewise penalty function whose breakpoints will depend on the composition assigned to each train service. This is because the passengers admitted in each piece depend on the composition assigned, which is a variable in the problem.

**Robustness.** As mentioned before, robustness is introduced through composition changes and empty movements. When a composition change is performed, multiple failures can occur, forcing the train to be parked for a long time and causing an incident. The mechanical system used to perform a composition change is automatic, but it often fails and requires extra time to enact the change. Moreover, during composition changes, the brakes’ pneumatic circuit must be joined or separated depending on the performed operation. This is always a difficult and complicating issue, and human resources are required to perform it. Above all, composition change times are overestimated to account for the effects mentioned above and to try to introduce robustness into the system. Finally, if a malfunction has occurred it must be contained to avoid cascading effects. Containment of cascading effects is easier if the incident occurs during off-peak hours when more RS material is available at the depot station and other material can depart from the station to attend to as much of the demand as possible. In our model, this is treated by harshly penalizing composition changes during rush hours (see coefficient $\vartheta_{s,t}$, which depends on the station $s$ and the time period $t$).

Similarly, empty movements during rush hours complicate network operation because they use the same infrastructure as commercial train services. In addition, they obviously require human resources. Although human resources are always available at depot stations to perform composition changes and empty movements, it is better to keep these resources in the depot station to alleviate possible incidents during rush hours. Therefore, empty movements during rush hours are also heavily penalized. It is also better to avoid (if possible) empty movements to destination depot stations with time-dependent capacities (i.e., stations that are shared with different lines). This idea is represented by the $\theta_{s,s',t}$ coefficient, which penalizes empty movements between depot stations $s, s'$ within departure time period $t$.

Finally, the system is made more robust by assigning only one material type per line (i.e., for every train service operating the same line, the material must be equal). This constraint allows for all material on one line to be swapped between different train services at depot stations serving that line. Thus, there will be more opportunities to swap train services if an incident occurs, and the propagation of the incident can be mitigated easily.
Moreover, at shared depot stations, there could be as many material types as there are lines using the station, but it will be shown that some lines share the same material type and depot stations so that the lines can interchange RS material.

### 5.4 The Rolling Stock Model

In the RS model (RSM), the compositions of the train services are determined. Train units are usually formed by two locomotives and one carriage of the same type. Train units from different types are not compatible and cannot be mixed in the same composition. For the Madrid case study (Figure 4.1), a composition may be formed of one or two train units. Changes in composition will only take place in the initial or final depot stations of the train service. The distribution of train units in the depot stations will be the same at the beginning and end of the planning period.

Light maintenance requirements are not included in the model because they are met during off-peak hours. Once the RS has been assigned, the train routing has to be determined (Cadarso and Marín [34]). When train sequences are designed, the location of all RS material in the space-time network will be known. Thus, material that requires light maintenance is assigned to a sequence containing light maintenance opportunities during off-peak hours. For heavy maintenance, the fleet size is supposed to be large enough to take the RS material requiring heavy maintenance out of the rapid transit network.

We include capacity constraints in depot stations. We could also include capacity constraints for arcs of the network. However, the scheduling of train services has been properly designed in the timetable planning: the capacity of every arc is considered and adequate supervision of the infrastructure manager exists in this planning phase (headway times are also ensured). Thus, arc capacity constraints are not necessary because they are automatically matched. Likewise, train services must not be longer than station platforms: for each train service we consider the shortest platform in its path, so we only include one constraint per train service for the platform length requirement.

In our model, the relationships between the data and variables are considered within a directed space-time graph, $G(S, A)$, where $S$ is the set of stations and $A$ is the set of arcs. Each arc $a$ is defined by $(s, t, s', t')$, where $s$ and $s'$ are the origin and destination nodes, $t$ is the departure time, and $t'$ is the arrival time. That is, $t' = t + t_a$, where $t_a$ is the time to move from $s$ to $s'$. It is assumed that this time is known and fixed for each arc. This means that in the RSM, in which an arc is denoted by $a$, this may be understood as
\[ a = (s, s', t). \]

In this space-time graph, the train services are given. These services are a known sequence of arcs, with a known departure time from the first station. We define the train set as the pair \((\ell, t)\), where \(\ell\) is the train service number and \(t\) is the departure time period, implicit in the train service number.

The RSM arises as an extension of the model proposed in Marín and Cadarso [91]. In the model presented in this chapter, special attention is given to shunting in depot stations and robustness. The RSM mathematical formulation follows:

- **Sets:**
  - \(L(\ell)\): the set of train services. Each train service is characterized by an origin, a destination and a departure time.
  - \(T(t)\): the set of time periods.
  - \(S(s)\): the set of stations.
  - \(A(a)\): the set of arcs.
  - \(M(m)\): the set of train unit material types.
  - \(C(c)\): the set of compositions.
  - \(A_{l_{\ell}}\) = 1 if arc \(a\) is used by the train service \(\ell\), and = 0 otherwise.
  - \(S_{l_{\ell}}\) = 1 if station \(s\) has the minimum platform length in service \(\ell\), and = 0 otherwise.
  - \(SC\): the set of depot stations.
  - \(CT\): the set of count time.

- **Parameters:**
  - \(c_{m,c}\): operating cost per rolled kilometer of composition \(c\) of train unit type \(m\).
  - \(i_{c,m}\): investment cost of train unit type \(m\). This parameter may represent a leasing cost for material from other lines.
  - \(p_{3-4_{a,\ell}}\): penalty per passenger in excess between 3 and 4 \(pax/m^2\) in arc \(a\) and train service \(\ell\).
  - \(p_{4-10_{a,\ell}}\): penalty per passenger in excess between 4 and 10 \(pax/m^2\) in arc \(a\) and train service \(\ell\).
• $\theta_{s,s',t}$: penalty for empty movement between depot stations $s, s'$ with departure time period $t$.

• $\vartheta_{s,t}$: cost of composition change in station $s$ and in time period $t$.

• $\alpha_{\ell,s,t}$: = -1, if train service $\ell$ leaves from station $s$ in time period $t$; = 1, if train service $\ell$ arrives at station $s$ in time period $t$; = 0, otherwise.

• $\chi_m$: fleet size for train units of type $m$.

• $g_{a,\ell}$: expected passenger flow in arc $a$ used by train service $\ell$.

• $q^3_m$: passenger capacity (seating+standing) for the 3 pax/m configuration in train unit of type $m$.

• $q^4_m$: passenger capacity (seating+standing) for the 4 pax/m configuration in train unit of type $m$.

• $q^{10}_m$: passenger capacity (seating+standing) for the 10 pax/m configuration in train unit of type $m$. This number is large to avoid infeasible solutions.

• $cap_{s,t}$: time-dependent capacity of station $s$ in time period $t$.

• $l_m$: length of a train unit of type $m$.

• $pl_s$: the platform length for each station $s$.

• $o_c$: number of train units in composition $c$.

• $r_s$: the rotation time duration in station $s$.

• $e_s$: the junction time duration in station $s$.

• $d_s$: the separation time duration in station $s$.

• $t_i, t_f$: the initial and final times of the planning period.

• $Km_\ell$: the number of kilometers rolled by train service $\ell$.

• $et_{s,s'}$: empty movement time from station $s$ to station $s'$.

• $\beta_{\ell,t',t}$: = 1, if train service $\ell$ that departed during time period $t'$ is still rolling during time period $t$; = 0, otherwise.
• $\gamma_{s,t',t'}: = 1$, if a rotation that finishes at station $s$ during time period $t'$ was being performed during time period $t; = 0$, otherwise.

• $\mu_{s,t',t'}: = 1$, if a composition change that started at station $s$ during time period $t'$ is still performing during time period $t; = 0$, otherwise.

• $\xi_{s,s',t,t'}: = 1$, if an empty movement between stations $s$ and $s'$ that departed during time period $t'$ is still rolling during time period $t; = 0$, otherwise.

- Variables:
  • $x_{\ell,m,c}: = 1$, if train service $\ell$ uses train unit type and composition $(m,c); = 0$, otherwise.
  
  • $em_{m,c,s,s',t}: = 1$, if empty movement from $s$ to $s'$ begins at station $s$ during period $t$ with train unit type and composition $(m,c); = 0$, otherwise.
  
  • $y_{m,c,s,t}:$ an integer variable, the number of compositions $c$ of type $m$, in station $s$ during time period $t$ (train inventory in station $s$).
  
  • $ym_m$: an integer variable, the number of train unit type $m$ to buy. Used to avoid infeasibilities in the model. It may also represent leasing costs from other lines.
  
  • $\pi_{3-4}^{a,\ell}: a$ positive variable, the number of passengers in excess between 3 and 4 $pax/m^2$ that use the train service $\ell$ at arc $a$.
  
  • $\pi_{4-10}^{a,\ell}: a$ positive variable, the number of passengers in excess between 4 and 10 $pax/m^2$ that use the train set $\ell$ at arc $a$.
  
  • $cc_{m,c,s,t}:$ an integer variable that counts composition changes performed at station $s$ in period $t$ from type and composition $(m,c)$.
  
  • $\epsilon_{m,c,s,t}: = 1$, if we begin the junction at station $s$ in period $t$ with type and composition $(m,c); = 0$, otherwise.
  
  • $\delta_{m,c,s,t}: = 1$, if we begin the separation at station $s$ in period $t$ with type and composition $(m,c); = 0$, otherwise.
  
  • $\rho_{m,c,s,t}: = 1$, if a rotation is finished at station $s$ in period $t$ with type and composition $(m,c); = 0$, otherwise.
The RSM for rapid transit networks is formulated as a multicommodity flow model. Some new aspects are contributed in the presented formulation: we have introduced multiple passenger capacities in the model depending on the configuration of standing passengers, using a piecewise formulation. In multicommodity flow balance constraints, we have included the material type and the composition, which can be understood as a subtype; composition changes have been included, which interchange commodities of different subtypes; similarly, rotations and empty movements have been included in these constraints. Moreover, a new group of constraints has been included to ensure that departing material has performed the mandatory shunting operations.

\[ \min z = \sum_{\ell \in L} \sum_{m \in M} \sum_{c \in C} c_{m,c} \cdot k_{m,\ell} \cdot x_{\ell,m,c} + \sum_{s,s' \in Sc} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} \theta_{s,s',t} \cdot c_{m,c} \cdot k_{m,s,s'} \cdot e_{m,s,s',t} + \sum_{s \in Sc} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} \varphi_{s,t} \cdot c_{m,c} + \sum_{m \in M} i_{m} \cdot y_{m} + \sum_{\ell \in L} \sum_{a \in Al_{\ell}} p_{a,\ell}^{3-4} \cdot \pi_{a,\ell}^{3-4} + \sum_{\ell \in L} \sum_{a \in Al_{\ell}} p_{a,\ell}^{4-10} \cdot \pi_{a,\ell}^{4-10} \]

In the objective function, a number of different costs are minimized. First, the operating costs of commercial train services are minimized. In the second term, the operating costs of empty movements are also minimized (cost per rolled kilometer equal to that of commercial service trains). However, the coefficient \( \theta_{s,s',t} \) increases operating costs for some empty movements to introduce robustness to the system. Another shunting cost, related to composition changes, is minimized in the third term. Through coefficient \( \varphi_{s,t} \), the cost of composition change is made dependent on station and time period. Robustness is introduced by minimizing composition changes, as these changes usually malfunction. A special cost is introduced in the fourth term to account for the possibility of leasing material from other lines. For computational purposes, this is equivalent to an infinite cost to avoid infeasibilities in the model. Finally, costs related to passengers in excess are introduced. The first cost appears if the density of standing passengers is between 3 pax/m² and 4 pax/m². Another cost is then introduced for standing passengers between 4 pax/m² and 10 pax/m². Both terms contribute to minimize the number of excess passengers.

Decision variables are subject to the constraints described in the following subsections.
5.4.2 Train Service Constraints

\[
\sum_{m \in M} \sum_{c \in C} x_{\ell,m,c} = 1 \quad \forall \ell \in L
\]

\[
\sum_{m \in M} \sum_{c \in C} o_c \cdot l_m \cdot x_{\ell,m,c} \leq p_{l,s} \quad \forall \ell \in L, s \in S_{l,\ell}
\]

Constraints (5.1) require that every train service is assigned a RS composition. Constraints (5.2) ensure that every train service matches the platform length requirements for its path.

5.4.3 Demand Constraints

\[
\sum_{m \in M} \sum_{c \in C} o_c \cdot q_m^3 \cdot x_{\ell,m,c} \geq g_{a,\ell} - \pi_{a,\ell}^{3-4} - \pi_{a,\ell}^{4-10} \quad \forall \ell \in L, a \in A_{l,\ell}
\]

\[
\pi_{a,\ell}^{3-4} \leq \sum_{m \in M} \sum_{c \in C} o_c \cdot (q_m^4 - q_m^3) \cdot x_{\ell,m,c} \quad \forall \ell \in L, a \in A_{l,\ell}
\]

\[
\pi_{a,\ell}^{4-10} \leq \sum_{m \in M} \sum_{c \in C} o_c \cdot (q_m^{10} - q_m^4) \cdot x_{\ell,m,c} \quad \forall \ell \in L, a \in A_{l,\ell}
\]

Constraints (5.3) require that the capacity assigned to each train service is sufficient to satisfy the passenger demand requirements. If the capacity is insufficient, the number of excess passengers is calculated. These passengers are limited in number by constraints (5.4) and (5.5), one for each group of passengers.

5.4.4 Material Constraints

\[
y_{s,t}^{m,c} + \sum_{\ell \in L} x_{\ell,m,c} + o_{c-1} \cdot \epsilon_{s,t-\ell}^{m,c-1} +
\]

\[
o_{c+1} \cdot \delta_{s,t-d_\ell}^{m,c+1} + p_{s,t}^{m,c} + \sum_{\ell' \in SC} \epsilon_{s',t-\ell'}^{m,c} =
\]

\[
y_{s,t}^{m,c} + \sum_{\ell \in L} x_{\ell,m,c} + o_{c+1} \cdot \epsilon_{s,t}^{m,c} +
\]

\[
o_{c-1} \cdot \delta_{s,t}^{m,c} + p_{s,t+r_{s}}^{m,c} + \sum_{\ell' \in SC} \epsilon_{s',t}^{m,c} \quad \forall s \in SC, t \in T,m,c \in M,C
\]
\[ \sum_{s \in S} \sum_{c \in C} o_c \cdot y_{t,m,c}^{m,c} + \]
\[ \sum_{t \in T} \sum_{c \in C} \beta_{t,t'}, t \cdot x_{t,m,c} + \]
\[ \sum_{s \in S} \sum_{c \in C} o_c \cdot \gamma_{s',t', t} \cdot \rho_{s,t'}^{m,c} + \]
\[ \sum_{s,s' \in S} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \gamma_{s',t', t} \cdot \rho_{s,t'}^{m,c} + \]
\[ \sum_{s \in S} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \xi_{s,s',t', t} \cdot e\epsilon_{s,s',t'}^{m,c} \]
\[ \leq \chi_m + y_{m,m} \quad \forall m \in M, t \in CT \]
\[ (5.7) \]

Constraints (5.6) describe the balance of train unit flow in each depot station for every time period. Train units parked in the immediately preceding period plus the train units arriving by commercial train services, empty movements and finished composition changes and rotations must be equal to the train units parked in the next period plus the departing commercial train services, empty movements and composition changes and rotations that begin in the next period. Constraints (5.7) require that the fleet size is large enough to satisfy the network flows. This is only verified at one period because of the material flow constraints.

### 5.4.5 Shunting Constraints

\[ \sum_{t \in T} \sum_{t' \in T} \sum_{m \in M, C} o_c \cdot y_{s,t,m,c}^{m,c} \]
\[ \sum_{t' \in T} \sum_{m \in M, C} \mu_{s',t', t} \left( o_{c+1} \cdot \epsilon_{s',t'}^{m,c} + o_c \cdot \delta\epsilon_{s',t'}^{m,c} \right) \]
\[ \sum_{s \in S} \sum_{t \in T} \sum_{m \in M, C} o_c \cdot \gamma_{s',t', t} \cdot \rho_{s,t'}^{m,c} \leq \text{cap}_{s,t} \quad \forall s \in S, t \in T \]
\[ (5.9) \]
\[ cc_{s,t}^{m,c} = \epsilon_{s,t}^{m,c} + \delta\epsilon_{s,t}^{m,c} \quad \forall s \in S, t \in T, m, c \in M, C \]
\[ (5.10) \]
\[ y_{t,m,c}^{m,c} = \rho_{t,m,c}^{m,c} \quad \forall s \in S, m \in M, c \in C \]
\[ (5.11) \]

Constraints (5.8) require that every departing commercial train service has performed the necessary rotation. Composition changes are included because, as stated above, they
also include the rotation time. Constraints (5.9) are depot station capacity constraints. The number of train units in each depot station is accounted for in every period to avoid exceeding the capacity \( cap_{s,t} \), which depends on time. Constraints (5.10) count every composition change into one single variable. Finally, constraints (5.11) ensure that the distribution of the fleet throughout the depot stations is equal at the beginning and end of the planning period.

5.4.6 Variable Domain

\[
x_{\ell,m,c} \in \{0, 1\} \quad \forall \ell \in L, m \in M, c \in C
\]

\[
e_{m,c}^{s,s',t} \in \{0, 1\} \quad \forall s, s' \in SC, t \in T, m \in M, c \in C
\]

\[
\epsilon_{s,t}^{m,c} \in \{0, 1\} \quad \forall s \in SC, t \in T, m \in M, c \in C
\]

\[
\delta_{s,t}^{m,c} \in \{0, 1\} \quad \forall s \in SC, t \in T, m \in M, c \in C
\]

\[
cc_{s,t}^{m,c} \in \mathbb{Z}^+ \quad \forall s \in SC, t \in T, m \in M, c \in C
\]

\[
y_{t}^{m,c} \in \mathbb{Z}^+ \quad \forall s \in SC, t \in T, m \in M, c \in C
\]

\[
y_{m}^{n} \in \mathbb{Z}^+ \quad \forall m \in M
\]

\[
\rho_{s,t}^{m,c} \in \mathbb{Z}^+ \quad \forall s \in SC, t \in T, m \in M, c \in C
\]

\[
n_{a,\ell}^{3-4} \in \mathbb{R}^+ \quad \forall \ell \in L, a \in Al_{\ell}
\]

\[
n_{a,\ell}^{4-10} \in \mathbb{R}^+ \quad \forall \ell \in L, a \in Al_{\ell}
\]

In the solution approach, some variable domains can be relaxed. Continuity in the excess passenger variable is justified because the passenger flow value is an expectation and could be non-integer. Moreover, some variables in the model formulation are defined as a sum of binary and integer variables, and may therefore be relaxed. These variables are \( y_{t}^{m,c} \), \( cc_{s,t}^{m,c} \), \( \rho_{s,t}^{m,c} \) and \( y_{m}^{n} \). Thus, constraints (5.16)-(5.17)-(5.18)-(5.19) can be replaced by constraints (5.22)-(5.23)-(5.24)-(5.25), respectively.

\[
cc_{s,t}^{m,c} \in \mathbb{R}^+ \quad \forall s \in SC, t \in T, m \in M, c \in C
\]

\[
y_{t}^{m,c} \in \mathbb{R}^+ \quad \forall s \in SC, t \in T, m \in M, c \in C
\]

\[
y_{m}^{n} \in \mathbb{R}^+ \quad \forall m \in M
\]

\[
\rho_{s,t}^{m,c} \in \mathbb{R}^+ \quad \forall s \in SC, t \in T, m \in M, c \in C
\]
5.5 Computational Experiments

All of our computational experience is for realistic cases drawn from RENFE’s regional network in Madrid, also known as "Cercanías Madrid" (Figure 4.1).

The network presented in this study case (Figure 4.1) is characterized by its modular structure. That is, in real life it is separated into different and independent modules for operating purposes. Every module has its own infrastructure (stations, depot stations, arcs, etc.). In this way, the RS material cannot be easily transferred from one module to another; it can be done, but extra human resources and time are needed to accomplish it; these extra resources are represented by the leasing cost in the objective function.

As mentioned above, some depot stations are shared between different lines. In these cases, we have used the total capacity of the station, and the model decides how to allocate the capacity to the material of different lines. However, some depot stations are shared between different modules. This is where the time-period-dependent capacity originates. The capacity assigned to each module is decided jointly with the infrastructure manager during Timetable planning, and the capacity assigned to each module may change during the planning period.

Two different cases (network modules) have been studied. The first case is line C5. This line can be considered an independent line for RS assignment purposes. However, it shares some depot stations. We have chosen this module because it has the highest frequency in the network. The second case consists of lines C3-C4. Although these are two different lines, they use the same material and share some depot stations; therefore, the two lines can interchange RS material.

Our runs were performed on a Personal Computer with an Intel Core2 Quad Q9950 CPU at 2.83 GHz and 8 GB of RAM, running under Windows Vista 64Bit, and our programs were implemented in GAMS/Cplex 11.1.

5.5.1 Study Case 1: Line C5

Line C5 has more than 320 train services scheduled each day with frequencies on the order of 3 minutes at rush hour, equivalent to the rotation time in this line. The line has 22 stations (Figure 5.1) and 4 depot stations: Mostoles el Soto, Atocha, Fuenlabrada and
Table 5.1 Train unit capacity and length

<table>
<thead>
<tr>
<th>Material Type</th>
<th>Seats</th>
<th>Standing</th>
<th>Density (Pax/m$^2$)</th>
<th>Length (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>240</td>
<td>261</td>
<td>3</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>348</td>
<td>4</td>
<td>870</td>
</tr>
<tr>
<td></td>
<td></td>
<td>870</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2 Line C5 RS model: the number of variables, constraints and non-zeros

<table>
<thead>
<tr>
<th></th>
<th>RSM</th>
<th>Reduced RSM</th>
</tr>
</thead>
<tbody>
<tr>
<td># Discrete variables</td>
<td>58628</td>
<td>47207</td>
</tr>
<tr>
<td># Continuous variables</td>
<td>175924</td>
<td>11575</td>
</tr>
<tr>
<td># Constraints</td>
<td>89317</td>
<td>19058</td>
</tr>
<tr>
<td># Non-zero elements</td>
<td>797321</td>
<td>139388</td>
</tr>
</tbody>
</table>

Humans. There is one material type available, and the train services can be of simple (one train unit) or double (two train units) composition.

The train units for the material in Line C5 have definite characteristics, as shown in Table 5.1. The train capacity is divided into seated and standing passengers. The number of seats is fixed, and a density value is defined for standing passengers. In every train unit, for example, for a density of 3 $\text{Pax/m}^2$ we would have 240 fixed seats plus 261 standing passengers (i.e., 501 passengers per train unit). However, for a density of 4 $\text{Pax/m}^2$, we would have the same 240 seats but a total capacity of 588 passengers.

In a daily planning period from 5:00 a.m. to 1:00 a.m. divided into one minute periods, we have 1200 time periods. The rotation time is 3 minutes at every depot station.

The RS MIP model size for this case is shown in Table 5.2. The RSM numbers of discrete and continuous variables, constraints and non-zero elements are given for the complete model (RSM) and for the reduced model (Reduced RSM) obtained using the Cplex presolver.
The primary model parameters for the operator are those penalizing excess passengers, \( (p_{a,t}^{3-4}, p_{a,t}^{4-10}) \) and the robustness parameters, \( (\theta_{s,s',t}, \vartheta_{s,t}) \).

Robustness parameters are obtained from operators. For example, in line C5, the Atocha depot station is shared among more than 5 different lines. This causes the capacity of C5 material to vary strongly during the planning period. Robustness parameters are chosen to try to avoid (if possible) composition changes and empty movements with the destination Atocha.

Once the robustness criteria are fixed, different solutions can be obtained by varying the penalties for excess passengers. These results are summarized in Table 5.3. For simplicity, we show three different cases in which the penalties for excess passengers are constant. The pair of excess passenger penalties \( (p_{a,t}^{3-4}, p_{a,t}^{4-10}) \) is listed in the first column, the number of train units used in the proposed solution \( (#C) \) in the second column, the train service operating costs (TSOC) in the third column, the empty movement operating costs (EMOC) in the fourth column, excess passenger costs (PEC) are shown in the fifth column, the number of composition changes \( (#CC) \) in the sixth column, the occupation index (OI) in the seventh column (obtained as an average value of every train service’s OI), and the solver solution time (ST) in seconds in the final column.

An important cost to the operator is the maintenance cost. For example, for this material, there is a daily fixed cost of, say, 400 €. Given the importance of the number of train units used in the network, different solutions must account for that number. If we compare any of the proposed solutions in Table 5.3 with the current solution provided by RENFE’s operators in Table 5.4, we can see that the number of used train units is slightly smaller. For different excess passenger penalties, we can see how the number of train units used varies; this variation is due to the number of composition changes performed, to the fact that increased penalties increase the PEC and more capacity must be offered, to material flows in the network subject to capacities, etc. For example, in Table 5.3 we can see how the solution with more composition changes is the solution that uses fewer train units.

Train service costs and empty movement costs are also reduced. However, excess passenger costs increase. This increase arises because if fewer train units are used, less capacity is offered to the passengers. For this reason, the occupation index also increases, and composition changes appear to enable different compositions of the train services. It is also worth noting that the solution time is nearly real time, substantially reducing the manual planning time.

As Atocha station is shared with many other lines and materials, there is no compo-
Table 5.3 Line C5 RS model solutions

<table>
<thead>
<tr>
<th>$p_{a,t}^{3-4},p_{a,t}^{4-10}$</th>
<th>#C</th>
<th>TSOC</th>
<th>EMOC</th>
<th>PEC</th>
<th>#CC</th>
<th>OI</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>66</td>
<td>78331.44</td>
<td>1635.76</td>
<td>3019</td>
<td>16</td>
<td>43.35</td>
<td>13.05</td>
</tr>
<tr>
<td>1.5</td>
<td>64</td>
<td>80099.76</td>
<td>1265.04</td>
<td>3554</td>
<td>20</td>
<td>42.30</td>
<td>24.56</td>
</tr>
<tr>
<td>3.6</td>
<td>65</td>
<td>82896.24</td>
<td>1281.36</td>
<td>4403</td>
<td>18</td>
<td>40.73</td>
<td>8.55</td>
</tr>
</tbody>
</table>

Table 5.4 Line C5 RS current solution operated by RENFE

<table>
<thead>
<tr>
<th>#C</th>
<th>TSOC</th>
<th>EMOC</th>
<th>PEC</th>
<th>#CC</th>
<th>OI</th>
</tr>
</thead>
<tbody>
<tr>
<td>74</td>
<td>109765.20</td>
<td>2232.12</td>
<td>874</td>
<td>0</td>
<td>28.30</td>
</tr>
</tbody>
</table>

sition change or empty movement with a destination of Atocha in any of the proposed solutions. However, empty movements from Atocha to other depot stations appear to provide capacity to this congested station.

As we have stated above, robustness may be introduced with different approaches. Some of these approaches could prevent dangerous empty movements and composition changes. Dangerous empty movements and composition changes are recognized as those occurring during rush hours. To illustrate this effect, two different cases are shown. In the first case, no robustness is introduced (NoRob) (i.e., $\theta_{s,s',t} = 1$ and $\vartheta_{s,t}$ is equal to its nominal value for all possible cases). In the second case, robustness is introduced (Rob), increasing the values of these parameters at rush hours from 7:00 a.m. to 10:00 a.m. and from 14:00 p.m. to 17:00 p.m. The computational results are shown in Table 5.5. In both cases, the rest of the parameters are those used in the previous examples.

We can see the differences between both cases in Table 5.5. In the second column, the train service operating costs (TSOC) are shown. In the robust case, these costs are greater because more train services are set to double composition; this cost may represent the robustness cost. In the third column, the empty movement costs (EMOC) are shown, which stay equal to those of the no robustness case. If we pay attention to empty movements during rush hours (#EMRH), we can see that the number of empty movements is reduced, achieving one of our goals. In a similar way, robustness is introduced in the solution by avoiding composition changes during rush hours (#CCRH). We note that for the cases studied, the passengers in excess costs (PEC) always remained similar or equal to those of the non-robust case. This is a desirable thing, because the price of robustness does not arise from passengers.
Table 5.5 Comparing non-robust and robust solutions for line C5

<table>
<thead>
<tr>
<th>Case</th>
<th>TSOC</th>
<th>EMOC</th>
<th>#EMRH</th>
<th>PEC</th>
<th>#CC</th>
<th>#CCRH</th>
<th>OI</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoRob</td>
<td>80099.76</td>
<td>1265.04</td>
<td>13</td>
<td>3554</td>
<td>20</td>
<td>4</td>
<td>42.30</td>
<td>24.56</td>
</tr>
<tr>
<td>Rob</td>
<td>80413.68</td>
<td>1265.04</td>
<td>10</td>
<td>3248</td>
<td>20</td>
<td>3</td>
<td>42.11</td>
<td>28.54</td>
</tr>
</tbody>
</table>

5.5.2 Study Case 2: Lines C3-C4

In lines C3-C4, there are nearly 400 scheduled train services each day with frequencies on the order of 10 minutes. In these lines, the rotation time is 2 minutes greater than that of line C5.

Line C3 is composed of 12 stations (Figure 5.2) and 3 depot stations: Chamartín, Atocha and Aranjuez. In line C4, there are 18 stations (Figure 5.3) and 8 depot stations: Parla, Parla Industrial, Getafe Centro, Atocha, Chamartín, Tres Cantos, Alcobendas and Colmenar Viejo. Some depot stations are shared between these lines.

There is one material type available, and the train services can have simple (one train unit) or double (two train units) compositions. Because the same material is used for both lines, RS can be interchanged between them.

The train units on lines C3-C4 have specific characteristics that are shown in Table 5.6. As shown, the train capacity is divided into seated and standing passengers. In every train
unit we have a capacity of 277 seats, but the number of standing passengers depends on their density, for example, for a density of 3 \( \text{pax/m}^2 \) we would add 360 standing passengers (for a total of 637 passengers per train unit), and so on.

In this second instance, the RS MIP model sizes of the complete model (RSM) and the reduced model (Reduced RSM) calculated by the Cplex presolver are shown in Table 5.7. The model is larger than that of line C5.

As for line C5, different solutions can be obtained by varying the penalties for passengers in excess, \( (p_{a,\ell}^3, p_{a,\ell}^{4-10}) \). These results are summarized in Table 5.8, which is composed of the same rows and columns that appear in Table 5.3. As we increase the excess passenger penalties, we can see that the TSOC increases because more capacity must be offered to meet the same demand with a higher quality of service. However, these costs are slightly lower in every presented solution than in RENFE’s current solution (Table 5.10). PEC are greater than in RENFE’s current solution but, as we have shown before, this increment is mainly due to the fact that we are penalizing passengers between 3 and 3.5 \( \text{pax/m}^2 \). However, these passengers are more comfortable because they are under the desired limit density of 3.5 \( \text{pax/m}^2 \), and thus the quality of service is not degrading. The ST is greater than that used for line C5 because the complexity of the model is greater for this study case.

As we have stated above, robustness may be introduced through different approaches, including avoiding dangerous empty movements and composition changes at rush hours. To illustrate this effect, two different case studies are shown, as for line C5: one in which no robustness is introduced (NoRob) and a second in which robustness is introduced.
Table 5.8 Lines C3-C4 RS model solutions

<table>
<thead>
<tr>
<th>$p_{a,t}^3$</th>
<th>$p_{a,t}^4$</th>
<th>#C</th>
<th>TSOC</th>
<th>EMOC</th>
<th>PEC</th>
<th>#CC</th>
<th>OI</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5,3</td>
<td>62</td>
<td>87170.89</td>
<td>3951.18</td>
<td>2751</td>
<td>28</td>
<td>26.60</td>
<td>171.92</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>63</td>
<td>87837.52</td>
<td>4090.83</td>
<td>2177</td>
<td>34</td>
<td>26.36</td>
<td>98.03</td>
<td></td>
</tr>
<tr>
<td>3,6</td>
<td>62</td>
<td>88094.56</td>
<td>4090.83</td>
<td>2462</td>
<td>34</td>
<td>26.27</td>
<td>142.60</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.9 Comparing the non-robust and robust solutions for lines C3-C4

<table>
<thead>
<tr>
<th>Case</th>
<th>TSOC</th>
<th>EMOC</th>
<th>#EMRH</th>
<th>PEC</th>
<th>#CC</th>
<th>#CCRH</th>
<th>OI</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoRob</td>
<td>87837.52</td>
<td>4090.83</td>
<td>20</td>
<td>2177</td>
<td>34</td>
<td>18</td>
<td>26.37</td>
<td>98.03</td>
</tr>
<tr>
<td>Rob</td>
<td>88085.11</td>
<td>4044.18</td>
<td>14</td>
<td>2177</td>
<td>32</td>
<td>16</td>
<td>26.34</td>
<td>204.48</td>
</tr>
</tbody>
</table>

(Rob). The computational results are shown in Table 5.9.

We can see the differences between the non-robust and robust cases in Table 5.9. For the robust case, TSOC are greater because more train services are set to double composition; this cost may represent the robustness cost. In the robust case, EMOC are lowered because robustness is achieved by penalizing them. If we pay attention to empty movements during rush hours (#EMRH), we can see that the number of empty movements is strongly reduced. In a similar way, the robustness introduced by composition changes is shown in columns #CC and #CCRH, representing the number of composition changes in the planning period and the composition changes during rush hours, respectively.

Again, PEC always remained similar or equal to the non-robust case. However, the ST was always greater in the robust case.

In Table 5.10 we compare the obtained solutions with the actual solution provided by RENFE. As can be seen from the results, the proposed solutions in Tables 5.8 and 5.9 improve the current solution, and even after introducing robustness into the model, costs are notably lowered.

5.6 Summary

The proposed mathematical model allows us to find solutions in a matter of minutes. The results are satisfactory, because in addition to commercial train services, they also account

Table 5.10 Current RS solution operated by RENFE for lines C3 and C4

<table>
<thead>
<tr>
<th>TSOC</th>
<th>EMOC</th>
<th>PEC</th>
<th>#CC</th>
<th>OI</th>
</tr>
</thead>
<tbody>
<tr>
<td>136633.86</td>
<td>6083.88</td>
<td>1550</td>
<td>0</td>
<td>15.20</td>
</tr>
</tbody>
</table>
for empty movements, adequate allocation of material in the depots and the optimal train unit mix forming the trains.

Robustness is introduced into the model through a number of different approaches. First, composition changes are penalized depending on the depot stations and time periods. Similarly, empty movements are penalized because they use the same infrastructure as commercial train services.

The results obtained in the network tests were satisfactory: operating costs were lowered while a high level of service quality for passengers was maintained and robust plans for network operation were provided. The possibility of analyzing several scenarios rather than just one is considered quite useful: by varying penalties for excess passengers, we can obtain different solutions providing different qualities of passenger service.
Chapter 6

ROBUST ROUTING OF RAPID TRANSIT ROLLING STOCK

The train routing problem determines the sequence for specific train units. In other words, once we know the train unit type assigned to each operation, we must know which operation precedes and succeeds it. We develop a robust model that attempts to minimize the delay propagation in each sequence as well as the crew requirements at depot stations. Here, robustness means that conflicting material connections are spread out in time as much as possible. Computational experiments are developed using the Madrid suburban rail network. The obtained results, achieved for short times and based on a previous efficient rolling stock assignment, show that a more robust and efficient solution than the current one can be obtained.

6.1 Introduction

The train routing problem (TRP) is the process of determining the best sequence for each train unit in the train network once the material assignment for each operation is known. The goal is to obtain sequences that minimize some cost to achieve a robust solution.

Train routing planning for suburban and urban operating railways must allow for each train to undergo different types of maintenance checks and robustness requirements, so the individual sequence of each train unit type is determined. Maintenance can be planned in detail in the sequences of individual train units. This method is usually applied to sparse systems. However, in high-density networks, plans might not be executed as planned because of disruptions in the operations. In our case, light maintenance is done during valley hours, and train units that require maintenance are assigned to sequences
with maintenance opportunities during valley hours, i.e., train units are swapped at the beginning of the planning period. We suppose that the fleet size is large enough to remove any train unit requiring heavy operation from the network.

Until now, RENFE planners have planned the RS assignment manually, which made the company’s operations rigid and inefficient. Thus, research has focused on investigating different possibilities to solve the RS problem and obtain robust and optimal solutions. However, this has led researchers to wonder how these solutions can be implemented in real life, because the RS problem does not identify which sequences each train unit must follow in the network. Thus, this chapter focuses on explaining an approach to solving the TRP and identifying the sequences to be assigned to each train unit.

This chapter is organized as follows: in Section 6.2, a literature overview is given. In Section 6.3, the TRP is described in detail. Section 6.4 introduces the mathematical formulation of the TRP. In Section 6.5, the mathematical model is tested in lines C3, C4 and C5 of the Madrid suburban network.

6.2 Literature Review

Historically, routing problems have been studied for the airline industry, and the primary objective was to find a feasible solution that ensured enough maintenance opportunities for each aircraft and maximized revenue. Clarke et al. [48] presented a routing model based on sequencing, discussed its similarity with the asymmetric traveling salesman problem with side constraints and solved the model using Lagrangian relaxation. Talluri and Gopalan [117] stated that it is possible to incorporate maintenance routing requirements into the fleet assignment model and formulate a joint model. However, the integer programming problem for such models becomes computationally intractable because the new maintenance constraints destroy the structure that makes it easy to solve the fleet assignment problem. Barnhart et al. [6] presented a string based model for fleet assignment and maintenance routing. Because the number of columns for this problem is huge, they exploited the network structure of the problem and presented a branch-and-price approach to solve it. Lan et al. [85] proposed a new approach to reducing delay propagation by intelligently routing aircraft using stochastically generated inputs. They presented computational results obtained using data from a major U.S. airline and showed that the delay propagation was significantly reduced.

There are different approaches to solve the TRP in railway planning. 1) One could build the sequences after the timetable is known, paying attention only to the timetable.
This is the case of Barnhart et al. [6], in which strings were built upon the timetable. 2) Another approach might be to build the sequences once the timetable and the RS assignment are known, thus, once the material type is assigned, one could also design maintenance opportunities, as in the cases of Clarke et al. [48] and Lan et al. [85]. 3) Last, RS assignment and TRP integration could be used; for example, Talluri and Gopalan [117] tried to integrate both phases.

There are two main options in the problem formulation in the second approach. One is based on the path formulation, in which paths are built externally to the model and then the optimal paths are chosen. Lan et al. [85] used this approach to select the sequences with the lowest delays. The second approach is based on sequencing, in which the model builds the paths. Clarke et al. [48] presented a model that builds sequences once the timetable and the material assignment have been done. The second approach is closer to the approach in this chapter. De Almeida et al. [57] studied the concept of robustness in railway production planning. The notion of robustness was introduced to evaluate the behavior of a schedule with regard to uncertainty. They state that robustness can be improved by reducing the propagation of delays and increasing the number of feasible resource allocation exchanges. They first proposed a model based on a path formulation, which is usually the case in the airline context. Then, they proposed a second model to build locomotive schedules, based on a flow formulation that was adapted to the optimization model developed in the R&D department of SNCF. Nielsen et al. [100] used a two-step model. Once the timetable is known, they first generate circulations, i.e., types of convoys are assigned to different trains, and then RS duties are generated to assign each convoy to a determined number of operations. They generated the duties externally to the model by identifying chains of operations performed by the same material type. Then they minimized some costs, such as carriage kilometers.

6.3 Train Routing Problem in Rapid Transit Networks

The TRP in this chapter has been studied for the rapid transit network introduced in 4.3. Prior to the train routing phase the rolling stock assignment must be known. Therefore, the rolling stock assignment in the previous chapter has been used as input. Then, the TRP is the process of determining the best sequence for each train unit in the network obtaining sequences that minimize some cost to achieve a robust solution.

There is a set of operations that is known. Among them we may cite train services
commercial trains operating in the network to meet passenger demand), empty movements and composition changes. All of them have determined schedules and a rolling stock assignment. The operations in the network are indexed by $i \in I$.

Two different approaches are used to introduce robustness to the problem: the first is to introduce penalties for delays in operations, and the second is to minimize the crew necessary to enable some technical requirements of the train units.

Delay penalization is one way to include robustness to the model. Every service is characterized by departure and arrival depot stations and times. However, between these depots there are other intermediate stops to satisfy passenger demand.

The demand in a planned situation may be defined by temporal arcs representing physical movements through stations with a determined departure time. Services are composed of several arcs, and the demand may vary strongly depending on departure time. Although the service time is fixed, it might change due to the time needed to embark and disembark passengers from the trains and other minor technical aspects, which is defined as propagated delay. We suppose that this time will increase during peak hours, delaying service arrival at the depot station, and that it will decrease during valley hours, advancing the time of arrival at the depot station.

As mentioned above, this delay may be positive or negative, and it will be a function of the service performed. Then, the delay in each operation will be propagated to the next operation in the sequence if the slack between both is not enough to absorb it.

In this way, the actual propagated delay from operation $i$ to operation $j$ is defined by $pd_{i,j} = \max(ad_i - slack_{i,j}, 0)$, where $ad_i$ is the arrival delay of operation $i$, and $slack_{i,j}$ is the planned slack between operations $i$ and $j$. When $pd_{i,j} > 0$, the train units performing operation $i$ will not be on time to perform operation $j$. However, they will be considered independent operations, and solutions will be penalized but feasible in the model formulation.

This independence is justified for RENFE performance, because in real operation, they do not permit propagated delays. Although there is a lack of capacity and resources in the network, RENFE planners always reserve some train units and, in case of delays, they swap to avoid delay propagation. The obtained statistical data might be considered independent of the sequence. Thus, this robustness criterion indirectly minimizes the number of necessary swapping operations and minimizes the human resources required to perform the material swapping.

In Figure 6.1, a sequence with three different operations is shown. Operation 1 has a planned timetable represented by the solid arrow. However, its timetable is changed
during its trip for some reason; the new timetable is indicated by the dotted arrow. Thus, operation 1 has an arrival delay, $ad_1$. Because $ad_1$ is greater than the available slack between operations 1 and 2, $slack_{1,2}$, the delay is propagated to operation 2. Furthermore, operation 2 will also be delayed because the number of passengers on the platforms will probably be increased. That is, a delay at a connection in the sequence produces a cascade effect in the network. However, by minimizing the expected delay for each connection, the perverse effects may be avoided. In real life, RENFE planners would have swapped some material, if possible, to perform operation 2 on time, preventing propagation of the delay.

![Figure 6.1 Sequence with three operations](image)

To determine the $ad_i$ for each operation $i \in I$, the delay distribution was obtained from the historical database of the RENFE network. Three months of data were collected. In order to prevent irregularities in the statistical study, days with significant incidents were removed from the study, but delays due to passengers embarking and disembarking were included.

We studied different possible probability distribution functions to represent $ad_i$, for example, normal, log-normal, gamma, etc. Because trains may arrive at the destination early, on time or late, we needed an asymmetric distribution.

The $\chi^2$ test and Kolmogorov test were used to determine if the $ad_i$ followed any of the proposed distributions. We found that the log-normal distribution fit the historical data for many stations over the entire day. With a significance level of 0.05, the null hypothesis was accepted for 90.6% of trains. Similar results have been obtained in the airline industry for the propagated airplane delay Lan et al. [85].

A variable $x$ is log-normally distributed if $y = \log(x)$ is normally distributed, with $\log$ denoting the natural logarithm. The general formula for the probability density function

63
is (NIST/SEMATECH e-Handbook of Statistical Methods [102]):

\[
f(x) = \frac{1}{(x - \theta)\sigma\sqrt{2\pi}} e^{-\frac{(\log(\frac{x}{m_i}))^2}{2\sigma^2}}
\]  

(6.1)

, where \(\sigma\) is the shape parameter, \(\theta\) is the location parameter and \(m\) is the scale parameter. The maximum likelihood estimate for the scale parameter, \(m\), for each operation \(i\) is \(m_i = e^{\mu_i}\), where \(\mu_i\) is the mean for operation \(i\).

To calculate the \(pd_{i,j}\) values, expressed in minutes, we take advantage of the fact that \(slack_{i,j}\) is always constant and can thus be captured through the location parameter as shown in (6.2). The propagated delay is calculated in (6.3) for each compatible connection pair \(i - j\):

\[
\theta'_{i,j} = \theta_i - slack_{i,j}
\]

\[
E[pd_{i,j}] = \theta'_{i,j} \left[ 1 - \phi \left( \frac{ln \left( \frac{a - \theta'_{i,j}}{m_i} \right)}{\sigma_i} \right) \right] + m_i \cdot e^{\mu_i} \left[ 1 - \phi \left( \frac{ln \left( \frac{a - \theta'_{i,j}}{m_i} \right)}{\sigma_i} - \sigma_i \right) \right]
\]

(6.3)

, where \(\phi(x)\) is the cumulative distribution function of a standard normal distribution, and \(a\) is equal to 0 if \(\theta'_{i,j} \leq 0\) or \(a\) is equal to \(\theta'_{i,j}\) otherwise.

The other way of introducing robustness may be through penalizing the crew requirement, \(cr_{i,j}\), due to brake malfunction between operations \(i\) and \(j\) because the train brakes are air compressed circuits and are emptied automatically when the train is stopped at a depot station. They must be inflated to have the trains ready to perform services. We could simulate this effect by assigning extra crews to depot stations, but this is an inefficient situation that can be avoided by adding a penalization term \(cr_{i,j}\), expressed in minutes in the objective function.

The time needed to inflate air compressed circuits varies with the amount of time that the train has been stopped. At first, we assume that it grows linearly with time, but maintains a constant value once they are totally empty. We propose the piecewise penalization function shown in Figure 6.2. For the rolling material used in line C5, the air compressed circuits are totally emptied in 10 minutes, and the required time to inflate them is 12 minutes; these times may vary with different material types and installations.

In Figure 6.1, we can again appreciate this effect. Assuming that operation 1 will be performed as planned, the connection time to operation 2 will be small enough to avoid
the necessity of extra crew resources. However, supposing that operation 2 will be also performed as planned, the connection time to operation 3 is very large, and extra human resources will be necessary to enable the material to perform the operation.

### 6.4 The Rapid Transit Routing Model

The solution obtained from the rapid transit rolling stock model identified the flow of different train unit types through the network. In this network, we had several commercial services to which we assigned a train. Some operations were necessary to make this assignment feasible, such as empty movements and composition changes.

The rapid transit rolling stock model model does not identify which specific train unit is assigned to each operation. Train routing is the process of assigning each individual train unit, referred to as an identification number, to operations. Given an assignment of train unit types to operations, we must determine a sequence of operations to be rolled by an individual train unit such that the assigned operations are included in exactly one sequence, and there is always the necessary number of train units available for every operation.

It is generally considered that the timetable will be executed as planned. However, the reality will probably be different than the planned situation. For this reason, expected delays are introduced into the model to try to avoid the delay propagation, making the system robust; we isolate the disruption effects that would cause a delay for the next train in the sequence by spreading out problematic connections. To minimize human resources, we include a cost that represents the crew requirement.

The Robust Rapid Transit Routing Model (RRTRM) minimizes the train delay and crew requirement subject to the continuity of feasible operations. The model is defined by the following items:
Sets:

$I(i)$: operations’ set, indexed by $i$ or $j$.

$T(t)$: periods’ set.

Parameters:

$co_{i,j}$: = 1, if operations $i$ and $j$ are compatible, that is, they have the same ending and starting stations and compatible compositions; = 0, otherwise.

$cr_{i,j}$: penalization by crew requirement.

$pd_{i,j}$: penalization by delay from operation $i$ to operation $j$.

$t_i$: period of time at which operation $i$ finishes.

$t_j$: period of time at which operation $j$ begins.

$\alpha_i$: = 2, if operation $i$ is an uncoupling operation; =1, otherwise.

$\beta_j$: = 2, if operation $j$ is a coupling operation; =1, otherwise.

$\gamma_{i,j}$: minimum time that a train unit needs to be available after operation $i$ and before operation $j$.

Variables:

$x_{i,j}$: = 1, if operation $i$ precedes operation $j$ in the sequence; = 0, otherwise.

The RRTRM mathematical formulation is based on sequencing:

\[
\min z = E \left[ \sum_{i \in I} \sum_{j \in I} pd_{i,j}x_{i,j} \right] + \sum_{i \in I} \sum_{j \in I} cr_{i,j}x_{i,j} \quad (6.4)
\]

Subject to:

\[
\sum_{j \in I} \sum_{\substack{t_i,t_j \in T \\text{such that} \ t_j \geq t_i + \gamma_{i,j}}} co_{i,j}x_{i,j} = \alpha_i \quad \forall i \in I \quad (6.5)
\]

\[
\sum_{i \in I} \sum_{\substack{t_i,t_j \in T \\text{such that} \ t_j \geq t_i + \gamma_{i,j}}} co_{i,j}x_{i,j} = \beta_j \quad \forall j \in I \quad (6.6)
\]

\[
x_{i,j} \in \{0,1\} \quad \forall i, j \in I \quad (6.7)
\]
The objective is to minimize the propagated delay and crew requirement. Constraints (6.5) establish that for each operation \( i \), another \( j \) follows it in the sequence, matching the composition and spatial-temporal requirements; if operation \( i \) is a disaggregation, two different operations must follow \( i \), so the summation must be equal to two. In a similar way, constraints (6.6) ensure that for each operation \( j \), another \( i \) precedes it in the sequence, matching composition and spatial-temporal requirements; if operation \( j \) is an aggregation, two different operations must precede \( j \), so the summation must be equal to two. Finally, constraints (6.7) define the variable domains.

In the network presented in Figure 4.1, the stations’ platforms have enough length to serve trains of at most two train units. In this way, the model has been developed under this assumption. However, this aspect may be easily adapted to situations with more than two train units by adjusting \( \alpha_i \) or \( \beta_j \) values and \( co_{i,j} \) value. For example, if we have a train composed of three train units and it must be disaggregated by operation \( i \) in two different trains (one composed of two train units and the other one by one train unit), we will have the following parameter values: \( \alpha_i = 3 \) and \( co_{i,j1} = 1 \), \( co_{i,j2} = 2 \); being \( j1, j2 \) operations composed of one train unit and two train units, respectively. Finally, an upper bound on the variable summation must be imposed in order to ensure that the train is decomposed in two parts and not in three. In a similar way, the study case will work with only one material type; however, the model formulation can deal with more material types. This is done by adjusting \( co_{i,j} \) value to 1 if the material is equal for both operations; 0, otherwise.

Our model is a stochastic discrete optimization problem without random variables in the constraints and with an expected value function that can be calculated easily, as explained above. Thus, the objective function can be rewritten as follows:

\[
\min z = E \left[ \sum_{i \in I} \sum_{j \in I} pd_{i,j} x_{i,j} \right] + \sum_{i \in I} \sum_{j \in I} cr_{i,j} x_{i,j} = \sum_{i \in I} \sum_{j \in I} (E[pd_{i,j}] + cr_{i,j}) x_{i,j} \quad (6.8)
\]

6.5 Case-Study: Madrid Suburban Rail Network

All computations in this chapter rely on "RENFE-Cercanías" Madrid network data (Figure 4.1). The RRTRM tests were done for lines C5 (Figure 5.1), C3 (Figure 5.2) and C4 (Figure 5.3) of the network.

The RRTRM was solved as a mixed-integer optimization problem using the branch and bound method. The model was implemented in GAMS/Cplex 11.1 on a PC with an
Table 6.1 Line C5 RRTRM model: number of variables, constraints and non-zeros

<table>
<thead>
<tr>
<th></th>
<th>RRTRM</th>
<th>Reduced RRTRM</th>
</tr>
</thead>
<tbody>
<tr>
<td># Discrete variables</td>
<td>17493</td>
<td>13729</td>
</tr>
<tr>
<td># Constraints</td>
<td>871</td>
<td>640</td>
</tr>
<tr>
<td># Non-zero elements</td>
<td>52090</td>
<td>27419</td>
</tr>
</tbody>
</table>

Intel Core2 Quad Q9950 CPU at 2.83 GHz and 8 GB of RAM, running under Windows Vista 64Bit.

6.5.1 Line C5: Computational Results

The RRTRM size for this instance is shown in Table 6.1. The numbers of discrete variables, constraints and non-zero elements are given for the complete model (RRTRM) and for the reduced model (Reduced RRTRM) obtained by the Cplex presolver.

In Tables 6.2 and 6.3 a brief summary of the optimal RRTRM solution is shown to illustrate its performance. Table 6.2 shows the optimal operation sequence for train unit 34 provided by the RRTRM. In the first column, we can see the number of operation. For line C5, we have a total number of 489 operations. Intermediate parking operations are not considered operations and are not shown in the table. In the second column, the performed services are shown. In the following columns we can see the origin (OS) and destination (AS) of the operations and the departure (DT) and arrival (AT) times for each one. Note that parking and composition change operations have the same origin and destination because they are performed in the same depot station. The \( pd_{i,j} \) column shows the penalization for delay between the operation in the same row and the one in the next row. The last column shows the penalization for crew requirements, \( cr_{i,j} \). When this penalization takes a value of 12 minutes, the air compressed circuits are totally emptied. The objective is to minimize the delay effects and crew requirement terms. Notice that convoy number 34 is only used during peak hours and is parked at depot stations for the rest of the day. This is due to the extra resources needed during peak hours, and consequently this extra resource is parked during valley hours.

Table 6.3 shows the sequence to be performed by train unit 55. In this case, the sequence begins in the simple composition. In this sequence, we can see an effect that produces a lack of capacity at the stations. In operation 342, we can see an empty movement to station 35006. Then, operation 397 is performed in that station, and the material goes back to 35012 in operation 372. Similarly, lack of capacity mandates operations 347 and 376 between the same previous stations.
Table 6.2 Sequence given by RRTRM for train unit 34

<table>
<thead>
<tr>
<th>Operation</th>
<th>Service</th>
<th>O S</th>
<th>D T</th>
<th>A S</th>
<th>A T</th>
<th>pd_{i,j}</th>
<th>cr_{i,j}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>35607</td>
<td>5:10</td>
<td></td>
<td>35607</td>
<td>6:42</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>19518</td>
<td>6:42</td>
<td>35012</td>
<td>7:41</td>
<td>4.0900</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Empty</td>
<td>35012</td>
<td>7:41</td>
<td>35002</td>
<td>7:45</td>
<td>3.9728</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>193</td>
<td>27341</td>
<td>7:48</td>
<td>35607</td>
<td>8:43</td>
<td>-</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>262</td>
<td>27410</td>
<td>15:03</td>
<td>35002</td>
<td>15:58</td>
<td>-</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>309</td>
<td>27457</td>
<td>19:42</td>
<td>35607</td>
<td>20:37</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Final Parking</td>
<td>35607</td>
<td>20:37</td>
<td>35607</td>
<td>24:45</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3 Sequence given by RRTRM for train unit 55

<table>
<thead>
<tr>
<th>Operation</th>
<th>Service</th>
<th>O S</th>
<th>D T</th>
<th>A S</th>
<th>A T</th>
<th>pd_{i,j}</th>
<th>cr_{i,j}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>35607</td>
<td>5:10</td>
<td></td>
<td>35607</td>
<td>5:15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>19502</td>
<td>5:15</td>
<td>35012</td>
<td>6:14</td>
<td>1.5962</td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td>Empty</td>
<td>35012</td>
<td>6:22</td>
<td>35002</td>
<td>6:26</td>
<td>0.6339</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>Aggregation</td>
<td>35002</td>
<td>6:35</td>
<td>35002</td>
<td>6:45</td>
<td>-</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>169</td>
<td>27317</td>
<td>6:48</td>
<td>35607</td>
<td>7:43</td>
<td>2.1935</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>19532</td>
<td>7:52</td>
<td>35012</td>
<td>8:51</td>
<td>1.5239</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Empty</td>
<td>35012</td>
<td>8:52</td>
<td>35006</td>
<td>9:18</td>
<td>-</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Disaggregation</td>
<td>35006</td>
<td>9:18</td>
<td>35006</td>
<td>9:28</td>
<td>-</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>372</td>
<td>35006</td>
<td>15:08</td>
<td>35012</td>
<td>15:11</td>
<td>1.2505</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>Empty</td>
<td>35012</td>
<td>15:15</td>
<td>35607</td>
<td>16:14</td>
<td>0.0774</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Aggregation</td>
<td>35006</td>
<td>16:39</td>
<td>35002</td>
<td>16:49</td>
<td>-</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>19612</td>
<td>17:27</td>
<td>35012</td>
<td>18:26</td>
<td>6.0836</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Empty</td>
<td>35012</td>
<td>18:26</td>
<td>35006</td>
<td>18:29</td>
<td>0.9615</td>
<td>8.4</td>
<td></td>
</tr>
<tr>
<td>Empty</td>
<td>35006</td>
<td>18:36</td>
<td>35012</td>
<td>18:39</td>
<td>0.3</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>376</td>
<td>35012</td>
<td>18:43</td>
<td>35002</td>
<td>19:42</td>
<td>4.002</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>446</td>
<td>35002</td>
<td>20:40</td>
<td>35002</td>
<td>24:45</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
In order to show that the obtained solutions are more robust than the current ones used by RENFE, a different objective function is proposed. Nowadays, RENFE planners design sequences manually without accounting for possible delays, although they do account for the time a convoy is stopped until the next operation in the sequence either due to brake malfunction if the stopped time is great or because their preferences are to keep a convoy parked until the next day once it has stopped. This operation can be represented as a minimization problem, minimizing the stopping times, subject to the same constraints in the RRTRM. In the remainder of the paper, the RENFE problem (RENFEP) refers to this minimization problem. The two different approaches are shown in Table 6.4. In the first column (RRTRM), the expected delay in minutes over an entire day for line C5 is shown. The second column (RENFEP) shows the expected delay obtained using the criteria of the RENFE planners. Finally, the reduction in the expected delay obtained using the RRTRM is shown in the last column. This expected delay reduction implies an induced efficiency; as said above, the delay propagation is usually avoided by swapping material. However, these operations require some reserve train units and crews. By reducing the expected delay, swapping operations are also minimized, thus improving the overall system efficiency.

Although the two problems, RRTRM and RENFEP, may look like opposite problems, with the first one minimizing some robustness cost and the second one minimizing the parking time, it should be noted that the RRTRM also accounts for parking times through the crew requirement term. Thus, the RRTRM solution is a compromise between the robustness and crew requirements.

Computational times are short enough to apply the model to real situations. For the case of line C5, the time taken to find the optimal RRTRM solution was 15.629 seconds. For the RENFEP, the computational time was slightly shorter, 15.167 seconds.

To illustrate that we achieved a robust solution with the RRTRM, we show three different sequences in the following figures. Two of them are feasible solutions but not optimal, and the third is the optimal solution given by the RRTRM. The commercial trains are presented in gray, the empty movements are illustrated as framed boxes, the composition changes are in black and the initial and final parkings are in white; notice that parking is considered an operation only for the parking time at the beginning and

<table>
<thead>
<tr>
<th></th>
<th>RRTRM</th>
<th>RENFEP</th>
<th>Expected Delay Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line C5</td>
<td>111.825</td>
<td>141.373</td>
<td>20.91 %</td>
</tr>
</tbody>
</table>
In Figures 6.3 and 6.4, the non-optimal but feasible solutions are illustrated. We can see the sequences for train units 34 and 55. The total delay penalization was obtained for each of them, adding the corresponding penalization for each operation connection. As we can see, the total penalization is high in both instances.

However, the operations in Tables 6.2 and 6.3, which compose the optimal solution provided by the RRTRM, are shown in Figure 6.5. The total delay penalization is much lower in this case than in the previous cases.

The proposed solution is based on an efficient RS circulation. Furthermore, we have routed the train units with attention to the expected delays and crew requirements. If we observe that the current solution operated by RENFE is not based on an efficient RS (Marín and Cadarso [91]) and that they route the material by only addressing feasibility (trying to account for parking times), we state that the obtained sequences are more robust and efficient than the actual sequences used.
Train unit 34
\[ \sum_{i,j \in I} p_{d_i,j} x_{i,j} = 8.062 \]
422 19 333 193 262 309 468

Train unit 55
\[ \sum_{i,j \in I} p_{d_i,j} x_{i,j} = 18.68 \]
412 3 330 379 169 33 342 397 372 84 383 110 347 376 113 302 446
401 373 ...
...80

Figure 6.5 Graphic representation of the optimal RRTRM solution for line C5

Table 6.5 Line C3-C4 RRTRM model: number of variables, constraints and non-zeros

<table>
<thead>
<tr>
<th></th>
<th>RRTRM</th>
<th>Reduced RRTRM</th>
</tr>
</thead>
<tbody>
<tr>
<td># Discrete variables</td>
<td>19582</td>
<td>14919</td>
</tr>
<tr>
<td># Constraints</td>
<td>1033</td>
<td>739</td>
</tr>
<tr>
<td># Non-zero elements</td>
<td>58302</td>
<td>29797</td>
</tr>
</tbody>
</table>

6.5.2 Lines C3 and C4: Computational Results

The RRTRM size for this case is shown in Table 6.5. The number of discrete variables, constraints and non-zero elements are given for the complete model (RRTRM) and for the reduced model (Reduced RRTRM) obtained by the Cplex presolver.

The effectiveness of the RRTRM was demonstrated through the computational results shown in Table 6.6. Here, the expected delay reduction is not as high, as line C5 is the line with the highest frequencies in the RENFE network, and a major improvement was consequently achievable in that case.

Computational times in this case were 24.024 seconds for the RRTRM and 23.162 seconds for the RENFEP. These times are greater than those obtained for line C5 as the model size is larger in this case.

Table 6.6 Lines C3-C4: expected delays in minutes

<table>
<thead>
<tr>
<th>Line C3-C4</th>
<th>RRTRM</th>
<th>RENFEP</th>
<th>Expected Delay Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>163.983</td>
<td>195.65</td>
<td>16.19 %</td>
</tr>
</tbody>
</table>
6.6 Summary

The robust train routing model developed in this chapter determines train units’ sequences that minimize the delay propagation and the crew requirements at depot stations.

The results obtained in the network tests are satisfactory: robust sequences of the train units are obtained; they are obtained near real time. We also minimize the crew resources necessary to perform all the sequences. However, further research is needed to generate complete crew planning.
Chapter 7

INTEGRATION OF TIMETABLE PLANNING AND ROLLING STOCK ASSIGNMENT IN RAPID TRANSIT NETWORKS

The aim of this chapter is to propose an integrated planning model to adequate the offered capacity and system frequencies to attend the increased passenger demand and traffic congestion around urban and suburban areas. In order to provide the problem more freedom to decide rolling stock flows and therefore better adjusting these flows to passenger demand, a new integrated model is proposed where frequencies are readjusted. The railway timetable and rolling stock assignment are also calculated. Difficult shunting operations will be penalized with the goal of selectively avoiding them and ameliorating their high malfunction probabilities. Swapping operations will also be ensured using homogeneous rolling stock material and ensuring parkings in strategic stations. We illustrate our model using computational experiments drawn from RENFE. The results show that through this integrated approach a greater robustness degree can be obtained.

7.1 Introduction

For any railway company, its efficiency in competitive markets depends strongly on its ability to define the railway timetabling so the demand meets a train at a desirable time. The railway company must provide the trains with the adequate material and composition so they have enough capacity to passengers feel comfortable in their trips.
The general aim of the Railway Timetabling (RT) problem is to provide a timetable for a number of trains on a certain part of the railway network based on the railway line planning frequencies. One may distinguish between cyclic and noncyclic timetables. A cyclic railway system is easy to remember for passengers because trains belonging to the same line always leave, for example, at \( x : 20 \) and \( x : 50 \). The noncyclic timetable is especially relevant on RTN, where the capacity of the infrastructure is limited. Then, the railway operators request to the infrastructure manager their preferred time slots.

The RT model makes decisions about trains departure times accounting for passenger demand during each time period. In an integrated approach, it also decides shunting schedules (empty movements and composition changes schedules). The RS model makes decisions about the RS assignment to trains and shunting in the depot stations. The integrated RT&RS problem can be stated as follows in the context of metropolitan RTN: given the expected numbers of passengers at each arc and during each period, and accounting for shunting operations’ schedule design, find the optimal RT and RS assignment.

The planned frequencies are known from the railway line planning problem. However, these planned frequencies were obtained without accounting for RS flows through the RTN. In order to provide the problem more freedom to decide RS flows jointly with train scheduling and therefore better adjusting these flows to passenger demand, we impose a minimum and maximum arc frequency values based on the previous obtained frequencies during the railway line planning. In this way, frequencies will be readjusted providing a more efficient network operation.

Major complicating issues are the available shunting capacities. Shunting is related to the need for RS to be parked in shunting areas when the RS is not used for traffic during off-peak hours and for those maneuvers to match compositions during times between the beginning and end of the planning period. These operations may sometimes be difficult to execute and they can easily malfunction, causing localized incidents that could propagate through the entire network due to cascading effects. In order to provide robustness to the problem, these operations will be penalized to selectively avoid them and their high malfunction probabilities.

Empty trains will also be moved to ensure RS availability because RS is a very limited resource during rush hours, when the passenger demand is very high. Another complicating issue is rotation times. Rotations are the maneuvers performed at depot stations to change the direction of motion of the RS.

This chapter presents a specialized problem describing RTN. The RT&RS Model (RT&RSM) will consider the optimization of trains’ departure times, rolling stock as-
ignment, empty trains and the optimal management of train units in depot stations, all while considering the character and capacities of these type networks.

This chapter is organized as follows. A literature overview is given in Section 7.2. We describe the problem for RTN in Section 7.3. In Section 7.4, the mathematical formulation is presented in detail. Section 7.5 contains the computational results based on a realistic case provided by RENFE.

7.2 Literature Review

Caprara et al. [45] provide an excellent state of art in railway optimization. They develop an introduction, a survey of available literature and a formal statement of the following problems: the RLP problem, the RT problem, the train platforming problem, the RS problem, the shunting problem and the crew planning problem.

The Railway Timetabling (RT) problem provides the train depart scheduling. There are two different possibilities when defining the RT: cyclic and non-cyclic timetables. Some references in cyclic RT are Nachtingall [98], Nachtingall and Voget [99] and Kroon and Peeters [83]; they develop the so-called Periodic Event Scheduling Problem (PESP), based on the paper of Serafini and Ukovich [112], a set of repetitive events is scheduled under cyclic time windows constraints. They do not consider the passenger demand. Gallo et al. [65] address a particular periodic scheduling problem arising when designing timetables in the railway freight transportation setting. They determine the departure times for each train and the route for each commodity to minimize the total delay time incurred by the commodities at the stations.

The non-cyclic RT is the common one for dense metropolitan networks. Carraresi et al. [46] consider the improvement of the effectiveness of the RTN, proposing a model which minimizes passengers’ total waiting time by modifying the departure time of the trips. They take as input a passenger assignment which satisfies the vehicles capacities. Brännlund et al. [25] discretize the time into time slots and subdivide the line into blocks. They use constraints in order to avoid the use of the same block during the same time slot by two different trains. Caprara et al. [43] define integer linear programming models based on graph representation. They solve it by Lagrangian relaxation to drive bounds in the context of a heuristic procedure. It produces good relaxations and heuristic solutions also for large size test problems. Caprara et al. [44] in the same context and methodology consider additional constraints: station capacities, prescribed timetable for a subset of trains, track out of order, etc. Cacchiani et al. [41] study the problem where both pas-
senger and freight trains are run. While the passenger trains have a prescribed timetable that cannot be changed, freight train operators send the infrastructure manager requests to insert new freight trains. The objective is to introduce as many new freight trains as possible by assigning them timetables.

Once the RT has been defined, the RS assignment must be done. An integer programming model is considered by Alfieri et al. [3] to determine the RS circulation for multiple RS types on a single line and on a single day. They use the concept of a transition graph to deal with this aspect. This concept is based on the assumption that for each trip, the next trip is known a priori. The model described by Alfieri et al. [3] was extended by Fioole et al. [63], to include combining and splitting trains, as happens at several locations in the Dutch timetable. They use an extended set of variables to locally obtain an improved description of the convex hull of the integer solutions. Robustness is considered by counting the number of composition changes. Maróti [94] focuses on planning problems that arise at NS (the main operator of passenger trains in the Netherlands). He identifies tactical, operational and short-term rolling stock planning problems and develops operations research models for describing them. The allocation of RS units to French TGV trains is studied by Ben-Khedher et al. [19]. The RS circulation must be adjusted to the latest demand known from the seat reservation system. Therefore, this problem contains a strong re-scheduling component. The objective is to maximize the expected profit for the company. A locomotive and carriage assignment problem was presented by Cordeau et al. [53]. The authors formulate the problem as a large integer program and use Benders decomposition to solve it. In a subsequent paper by Cordeau et al. [54], their model was extended by considering various aspects such as maintenance of the RS. A RS circulation problem related to the circulation of ICE train units in the German ICE network was described by Mellouli and Suhl [95]. In this case, the required capacities of the trains are known a priori. Cadarso and Marín [34] define a model to study suburban rapid transit RS with convoys formed by three cars of the same type. The trains may be composed of one or two convoys in a dense network to attend to asymmetric demand and scheduling.

Liebchen and Möhring [87] propose to integrate vehicle scheduling into the task of periodic timetabling. They demonstrate that the modeling capabilities of the PESP are not limited only to periodic timetabling. They use the well-established technique of minimizing the number of vehicles required to operate a periodic timetable by penalizing waiting times of vehicles. However, they do not assign specific vehicles to trains.
7.3 Timetable Planning and Rolling Stock Problem in Rapid Transit Networks

**Supply.** In previous chapters, train services were defined by $\ell \in L$. This set had information about the train service such as origin, destination and departure time. However, the definition of the set $L$ is slightly changed in this chapter. Let $L$ be the set of train lines. They are defined by an origin and a destination, both depot stations. A train line is a passenger train traveling from a depot station to another depot station stopping at a determined number of intermediate stations. The model decides whether a train line is assigned to a departure time or not, based on the knowledge of passenger demand in each arc and period.

Each train line $\ell \in L$ starts at depot station $SD_\ell \in SC$ and ends at depot station $ED_\ell \in SC$. For each infrastructure arc $a \in A$ we let $LA_a \subseteq L$ denote the set of lines that use arc $a$. The line $\ell$ is of length $km_\ell$ kilometers.

For safety purposes a minimum separation time is imposed between two consecutive trains using the same infrastructure, that is, the headway (in order to avoid congestion and possible incidents). Thus, the headway must be maintained everywhere. For the cases where a central station exists it will be enough to impose the headway in this station for arriving and departing trains. We mean by central station as a station through which every train line comes.

Each train line $\ell$ will have to be assigned a departure time $t \in T$ and a RS material $m \in M$ and composition $c \in C$.

**Passenger demand.** A known demand must be met by a given fleet. We assume that the forecasted travel demand is represented by a given origin destination matrix. However, this data is always difficult to obtain or/and to maintain it updated and does not represent passenger willings accurately. It is usually assumed that each passenger uses a pre-specified path through the network: each passenger is supposed to travel along the shortest-distance path through the network from his/her origin to his/her destination. In this way, with the available origin destination matrix and passenger counts in the network, we can obtain a passenger flow through arcs associated to time periods approximately.

The passenger demand for this problem is treated as a passenger flow $pf_{a,\tau}$ through each arc $a$ during time period $\tau \in TD_a$. Let $TD_a$ represent the time periods through which demand is counted in each arc $a$. We also include into the model the possibility of neglecting passengers from the system. That is, if passengers are not attended during the time periods they are willing to travel, they are denied. $\tau$ is a time interval containing
different time periods \( t \). That is, \( \tau \subset [t_n, t_{n+k}] \), where \( t_n \) represents any time period and \( k \) is a scalar to be defined. The scalar \( k \) measures the aggregation level of the passenger demand. That is, if \( k \) is small we are not aggregating passengers’ flows and they must be attended at their desired times or they are denied from the system. If it is greater, we let the model to aggregate the demand.

We assume that passenger demand waits in the station until a train comes. If there is enough capacity, every passenger will go on the train. Otherwise, passengers must wait until the next train coming in the station during time interval \( \tau \). If the train does not come, remaining passengers will be denied from the system.

Passenger flow is directly related with the timetable that was being used when the demand study was made. However, we use this data to change the timetable. First, we must realize that we are studying rapid transit systems where frequencies are very high, so passengers do not care so much the timetable but the frequency. In this way, according to the RLP problem we impose a minimum and maximum frequency values in determined arcs of the network. Hence, the new solution will not be so far from the previous one planning arc frequencies. Moreover, a minimum separation time is imposed for every train path departing from each depot station. In this way, the offered capacity, in terms of frequency, in the new solution will be similar to the previous one and enough sparse in time.

The operator will always try to offer comfortable capacity in each train line. It is obvious that when more capacity is offered, more passengers will travel comfortably. However, offering more capacity increases operating costs dramatically. Therefore, the composition assigned to each train line will be a tradeoff between the operating costs and the passengers’ comfort level.

**Robustness.** Railway planning is currently divided into several optimization steps from first strategic decisions to daily operations. It is well known that disintegrated planning produce optimal solutions for each stage but non-optimal global solutions. We integrate, so far separately considered, key planning stages (railway timetabling and rolling stock planning), and we carry on the research on integrating robustness into railway planning. That is, robustness is in part achieved by integrating different planning phases, because in our case the RS assignment and shunting operations depend strongly on RT, so a greater robustness degree can be achieved with the proposed combined approach.

The general meaning of the integration is as follows: consider two consecutive planning stages. In order to obtain a high-quality solution in the second stage, some slack must be introduced in the first stage. In the same way, a robust solution adds slack to safeguard
against data perturbation. This slack may provide a smooth interface between subsequent planning stages. This becomes a connection between robustness and integration.

Robustness is also introduced penalizing composition changes and empty movements. Infrastructure and human resources are needed to perform them. Although human resources are always available at depot stations to perform composition changes and empty movements, it is better to keep these resources in the depot station to alleviate possible incidents during rush hours. Containment of cascading effects is easier if the incident occurs during off-peak hours when more RS material is available at the depot station and other material can depart from the station to attend to as much of the demand as possible. In our model, this is treated by harshly penalizing composition changes and empty movements during rush hours.

Finally, the system is made more robust by assigning only one material type per line (i.e., for every train service operating the same line, the material must be equal). This constraint allows for all material on one line to be swapped between different train services at depot stations serving that line. Thus, there will be more opportunities to swap train services if an incident occurs, and the propagation of the incident can be mitigated easily. In order to ensure that at least one swapping operation is always possible at both extremes of train lines, we mandate the model to provide solutions with free parkings in depot stations at lines’ extremes. In this way, the obtained solution will be more resistant to possible delays because they may be alleviated by swapping trains.

7.4 The Timetable Planning and Rolling Stock Model

In this section the RT&RSM integrated formulation is presented. It is based on the formulation presented by Cadarso and Marín [34], where the RS assignment was done to a known schedule. Then, the main difference with respect to the cited work is that the timetable is a variable in the problem. That is, given train lines departure times must be determined. In order to match infrastructure manager requirements, headway constraints must be used to ensure a minimum separation time between trains using the same infrastructure. Moreover, frequency values are not fixed. For some arcs this value is upper and lower bounded in order to provide a desired service level. Another important difference is the demand treatment. Passenger flows associated to arcs and time periods are well known. Different demand aggregation levels may be studied with this new formulation depending on the value of the scalar $k$.

This model formulation is based on the assumption of that an unique mode of transport
exists. Thus, if the offered capacity is not enough, the demand will be neglected assuming that it will use another mode of transport. However, choosing the demand value according to the operator database we can assume that almost the whole demand will use the RTN under the above explained assumptions.

7.4.1 Sets

- \( S \) is the set of stations.
- \( SC \subseteq S \) denotes the set of depot stations.
- \( A \) is the set of arcs. The arcs represent the infrastructure. Each arc \( a \in A \) is from departure station \( ds_a \in S \) to arrival station \( as_a \in S \).
- \( L \) set of train lines defined by departure depot station, arrival depot station and every intermediate arc.
- \( LA_a \subset L \) is the set of train lines containing arc \( a \).
- \( T \) is the set of time periods. A time period \( t \in T \) represents a certain interval in time, for example from 8:00 to 8:01. \( \tau \) represents a time interval grouping \( k \) different time periods \( t \): \( \tau \subset [t_n, t_{n+k}] \), where \( t_n \) represents a determined time period.
- \( TD_a \subseteq T \) is the set of time intervals through which the demand is counted in each arc \( a \).
- \( M \) is the material set.
- \( C \) is the composition set.
- \( CT \subseteq T \) is the set of time intervals where the trains are counted.
- \( LM, LO \): set of train lines \( l \) going in one sense of the line; set of train lines going in the opposite sense of movement.
- \( LCS_{t,t} \): set of time periods during which train line \( \ell \) departs from the origin such that it is coming through central station during time period \( t \).
- \( LAT_{t,a,\tau} \): set of time periods during which train line \( \ell \) departs from its origin such that it is coming through arc \( a \) during \( \tau \).
- \( LA_s \): set of train lines arriving in depot station \( s \).
• \(LD_s\): set of train lines departing from depot station \(s\).

• \(AMF\): set of arcs \(a\) with a bounded frequency.

### 7.4.2 Cost parameters

We consider the following cost parameters in the objective function:

- \(c_{m,c}^{\ell,t}\) is the operating cost for train line \(\ell\) departing during time period \(t\) with material \(m \in M\) and composition \(c \in C\).
- \(ec_{m,c}\) is the empty movement operating cost per kilometer with material \(m \in M\) and composition \(c \in C\).
- \(\theta_{s,s',t}\) is a penalty value for empty movements between stations \(s, s'\) departing during time period \(t\).
- \(\vartheta_{s,t}\) is the cost of performing a composition change in depot station \(s\) during time period \(t\). This cost may be increased in order to penalize composition changes in some stations during certain time periods.
- \(epc_{a,\tau}^{3-4}\) is the penalty per passenger in excess between 3 and 4 \(pax/m^2\) in arc \(a\) during time period \(\tau \in TD_a\).
- \(epc_{a,\tau}^{4-up}\) is the penalty per passenger in excess above 4 \(pax/m^2\) in arc \(a\) during time period \(\tau \in TD_a\).
- \(upc_{a,\tau}\) is the penalty per unattended passenger in arc \(a\) during time period \(\tau\).
- \(ic_m\) is the investment or leasing cost for a train unit of material \(m \in M\).
- \(km_{s,s'}\) is the distance in kilometers from \(s\) to \(s'\).

### 7.4.3 Other parameters

- \(pf_{a,\tau}\) passenger flow through arc \(a\) during time period \(\tau\).
- \(q_{m,c}^{3-down}, q_{m,c}^{3-4}, q_{m,c}^{4-up}\) passenger capacity of material \(m \in M\) and composition \(c \in C\) for different capacities: below 3 \(pax/m^2\), between 3 and 4 \(pax/m^2\) and above 4 \(pax/m^2\), respectively.
• $\chi_m$ is the fleet size for each material $m$, i.e. The maximum number of train units that may run at the same time.

• $f_a$ minimum frequency in arc $a$.

• $\overline{f}_a$ maximum frequency in arc $a$.

• $tt_\ell$ trip time for train line $\ell$.

• $st$ separation time between two consecutive trains.

• $r_s$ rotation time in depot station $s$.

• $e_s$ coupling time in depot station $s$.

• $d_s$ uncoupling time in depot station $s$.

• $o_c$ number of train units in composition $c$.

• $t_i, t_f$ initial and ending time periods in the planning period.

• $\text{cap}_{s,t}$ capacity of depot station $s$ during time period $t$.

• $\beta_{\ell,t',t} = 1$, if train line $\ell$, which departed during time period $t'$, is still rolling during time period $t$; $=0$, otherwise.

• $\gamma_{s,t',t} = 1$, if a rotation performed in depot station $s$, which finishes during time period $t'$, was being performed during time period $t$; $=0$, otherwise.

• $\xi_{s,s',t',t} = 1$, if an empty movement performed between depot stations $s, s'$, which started during time period $t'$, is still being performed during time period $t$; $=0$, otherwise.

• $\mu_{s,t',t} = 1$, if a composition change performed in depot stations $s$, which started during time period $t'$, is still being performed during time period $t$; $=0$, otherwise.

7.4.4 Variables

The following variables are used in the model:

• $x_{m,c,\ell,t}^t$: binary variable. $=1$, if train line $\ell \in L$ is assigned to departure time $t \in T$ using composition $c \in C$ and material $m \in M$; $=0$, otherwise.
• \(em_{s,s',t}^{m,c}\): binary variable. =1, if an empty movement is performed between stations \(s\) and \(s'\) departing during time period \(t\) with material \(m\) and composition \(c\); =0, otherwise.

• \(\epsilon_{s,t}^{m,c}\): binary variable. =1, if a coupling operation is performed in station \(s\) starting during time period \(t\) with material \(m\) and composition \(c\); =0, otherwise.

• \(\delta \epsilon_{s,t}^{m,c}\): binary variable. =1, if an uncoupling operation is performed in station \(s\) starting during time period \(t\) with material \(m\) and composition \(c\); =0, otherwise.

• \(\rho_{s,t}^{m,c}\): binary variable. =1, if a rotation operation is performed in station \(s\) finishing during time period \(t\) with material \(m\) and composition \(c\); =0, otherwise.

• \(yt_{m,c}^{s,t}\): integer variable. It denotes the train inventory in depot station \(s\) during time period \(t\) of material and composition \(m,c\).

• \(h_{3-4}^{a,\tau}\): positive variable. It denotes the number of passengers in excess between 3 and 4 \(\text{pax/m}^2\) in arc \(a\) during time period \(\tau \in TD_a\).

• \(h_{4-up}^{a,\tau}\): positive variable. It denotes the number of passengers in excess above 4 \(\text{pax/m}^2\) in arc \(a\) during time period \(\tau \in TD_a\).

• \(dp_{a,\tau}\): positive variable. It denotes the number of disrupted passengers in in arc \(a\) during time period \(\tau\).

• \(ym_{m}\): integer variable. It denotes the number of train units of material \(m\) to buy or borrow from other lines.

In the following subsections we introduce the Timetable Planning and Rolling Stock Model (TTP&RSM) formulation.

### 7.4.5 Objective Function

\[
\min \sum_{t \in T} \sum_{s,s' \in SC} \sum_{m \in M} \sum_{c \in C} x_{m,c}^{s,s',t} + \sum_{s,s' \in SC} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} \theta_{s,s',t}^{c} ec_{m,c} km_{s,s',t}^{m,c} + \sum_{s \in SC} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} \theta_{s,t}^{m,c} \epsilon_{s,t}^{m,c} + \sum_{m \in M} i c_{m} y m_{m} + \sum_{a \in A} \sum_{\tau \in TD_a} e p c_{a,\tau}^{3-4} h_{a,\tau}^{3-4} + \sum_{a \in A} \sum_{\tau \in TD_a} e p c_{a,\tau}^{4-up} h_{a,\tau}^{4-up} + \sum_{a \in A} \sum_{\tau \in TD_a} u p c_{a,\tau} dp_{a,\tau}
\]

In the objective function the following terms have been considered, every one representing an economic penalization:
• operating cost for every line with a determined material and composition,
• operating costs for empty movements,
• cost of performing composition changes,
• investment or leasing cost in new material,
• penalty for passengers in excess of comfortable capacity and,
• penalty for unattended passengers.

7.4.6 Service constraints

$$\sum_{\ell} \sum_{t} \sum_{m} \sum_{c} x_{\ell,t}^{m,c} \geq f_a \quad \forall a \in AMF$$ (7.1)

$$\sum_{\ell} \sum_{t} \sum_{m} \sum_{c} x_{\ell,t}^{m,c} \leq f_a \quad \forall a \in AMF$$ (7.2)

$$\sum_{t \geq t_1} \sum_{\ell} \sum_{t_1} \sum_{m} \sum_{c} x_{\ell,t_1}^{m,c} \leq 1 \quad \forall t \in T$$ (7.3)

$$\sum_{t \geq t_1} \sum_{\ell} \sum_{t_2} \sum_{m} \sum_{c} x_{\ell,t_2}^{m,c} \leq 1 \quad \forall t \in T$$ (7.4)

Constraints (7.1) ensure that a minimum frequency is offered in some arcs. The maximum frequency value for the arcs is imposed by constraints (7.2). Constraints (7.3) ensure that the headway is maintained in one of the senses of movement of the RTN infrastructure, that is, we only account for the set of train lines going in that sense (LM). Similarly, constraints (7.4) ensure the headway for the set of train lines in the other sense of movement (LO). For each train line \( \ell \) and its possible time \( t_2 \) during which it is coming through central station we account for its departure time from the origin \( LCS_{t_2} \); then we add every train line coming through this station between a time instance \( t \) and the time instance \( t + st \) and we obey it to be equal or lower to 1.

7.4.7 Passenger constraints

$$\sum_{\ell} \sum_{t} \sum_{t_1} \sum_{m} \sum_{c} q_{m,c}^{3-down} x_{\ell,t_1}^{m,c} \geq$$

$$pf_{a,\tau} - h_{a,\tau}^{3-down} - h_{a,\tau}^{4-up} - dp_{a,\tau} \quad \forall a \in A, \tau \in TD_a$$ (7.5)
\[
\sum_{\ell \in L_{A}} \sum_{t \in LAT_{\ell,a,\tau}} \sum_{m \in M} \sum_{c \in C} \left( q_{m,c}^{3-4} - q_{m,c}^{3-down} \right) x_{m,c}^{\ell,t} \geq h_{a,\tau}^{3-4} \quad \forall a \in A, \tau \in TD_{a}
\]

\[
\sum_{\ell \in L_{A}} \sum_{t \in LAT_{\ell,a,\tau}} \sum_{m \in M} \sum_{c \in C} \left( q_{m,c}^{4-up} - q_{m,c}^{3-4} \right) x_{m,c}^{\ell,t} \geq h_{a,\tau}^{4-up} \quad \forall a \in A, \tau \in TD_{a}
\]

Passenger capacity constraints ensure that for each arc \( a \in A \) and each time interval \( \tau \in TD_{a} \), the capacity of the train lines is enough to accommodate the passenger demand \( pf_{a,\tau} \) minus the passengers that are in excess (denoted by variables \( h_{a,\tau}^{3-4}, h_{a,\tau}^{4-up} \)) or unattended (denoted by variable \( dp_{a,\tau} \)). The available capacity is the add of all possible train lines that will come through arc \( a \) and during each time interval \( \tau \). This information is in \( LAT_{\ell,a,\tau} \), where for each pair \((a, \tau)\) and train line \( \ell \), we have the departure time during which it departed from its origin.

Passenger in excess of comfortable capacity constraints ensure that passengers in excess are bounded by the different available capacities.

### 7.4.8 Rolling stock constraints

\[
y_{t}^{m,c} + \sum_{\ell \in LA_{s}} x_{m,c}^{\ell,t} + o_{c-1} \cdot \epsilon_{m,c}^{s-1} + o_{c+1} \cdot \epsilon_{m,c}^{s+1} + \sum_{s' \in SC} \epsilon_{m,c}^{s',s,t} - \sum_{s' \in SC} \epsilon_{m,c}^{s',s,t} =
\]

\[
y_{t}^{m,c} + \sum_{\ell \in LD_{s}} x_{m,c}^{\ell,t} + o_{c+1} \cdot \epsilon_{m,c}^{s+1} + \sum_{s' \in SC} \epsilon_{m,c}^{s',s,t} \quad \forall s \in SC, t \in T, m \in M, c \in C
\]

\[
\sum_{\ell \in LD_{s}} x_{m,c}^{\ell,t} = \rho_{s,t}^{m,c} + \epsilon_{m,c}^{s-1} + \delta_{s,t}^{m,c+1} \quad \forall s \in SC, t \in T, m \in M, c \in C
\]
\[
\begin{align*}
\sum_{s \in SC} \sum_{c \in C} o_c \cdot y^{m,c}_{s,t} + \\
\sum_{t \in T} \sum_{v \in T} \sum_{c \in C} o(c) \cdot \beta_{t,v} \cdot y^{m,c}_{t} + \\
\sum_{s \in SC} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \gamma_{s,v} \cdot \rho^{m,c}_{s,t} + \\
\sum_{s,s' \in SC} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \xi_{s,s',v} \cdot e m^{m,c}_{s,s',t'} \\
\sum_{s \in SC} \sum_{t' \in T} \sum_{c \in C} \mu_{s,t} \left( o_{c+1} \cdot \epsilon_{s,t'}^{m,c} + o_c \cdot \delta_{s,t'}^{m,c} \right) \\
\sum_{s \in SC} \sum_{t' \in T} \sum_{c \in C} y^{m,c}_{s,t} \leq \chi_m + y m \quad \forall m \in M, t \in C T \\
\sum_{m \in M} \sum_{c \in C} o_c \cdot y^{m,c}_{s,t} + \\
\sum_{t' \in T} \sum_{m \in M} \sum_{c \in C} \mu_{s,t',t} \left( o_{c+1} \cdot \epsilon_{s,t'}^{m,c} + o_c \cdot \delta_{s,t'}^{m,c} \right) + \\
\sum_{t' \in T} \sum_{m \in M} \sum_{c \in C} o_c \cdot \gamma_{s,t,t'} \cdot \rho^{m,c}_{s,t'} \leq cap_{s,t} \quad \forall s \in SC, t \in T \\
y^{m,c}_{s,t} = y^{m,c}_{s,t_f} \quad \forall s \in SC, m, c \in M, C
\end{align*}
\] 

Material and composition conservation constraints \(7.8\) ensure the convoys flow balance. These constraints ensure that the train number for every station and material at period \(t-1\), plus the arriving trains, minus the departing ones is equal to the train number at period \(t\).

Rotation and departure constraints \(7.9\) ensure that a rotation is performed before each train service departure.

In order to avoid that some of the time (i.e., \(t - t \ell, t - \ell s, t + r s\)) is outside set \(T\), some variables are fixed to 0 in the mathematical program. This situation affects to constraints \(7.8\) and \(7.9\).

Fleet capacity constraints \(7.10\) ensure that the number of trains used during time \(t \in T\) is limited by the size of the fleet. Note that these constraints count the running trains and those ones in depot stations. The variable \(yn_m\) specifies the additional trains to borrow from other lines at high cost.

Depot capacity constraints \(7.11\) ensure that station capacity is not overpassed. They account for parked and shunting material.

Trains initial and final constraints \(7.12\) denote that the number of convoys and compositions available in the stations at the beginning and end of the planning period are equal.
7.4.9 Variable Domain

\[ x_{t, m}^{c, m} \in \{0, 1\} \quad \forall \ell \in L, t \in T, m \in M, c \in C \]  
(7.13)

\[ m_{s, s'}^{c, m} \in \{0, 1\} \quad \forall s, s' \in SC, t \in T, m \in M, c \in C \]  
(7.14)

\[ c_{s, t}^{m, c} \in \{0, 1\} \quad \forall s \in SC, t \in T, m \in M, c \in C \]  
(7.15)

\[ \delta c_{s, t}^{m, c} \in \{0, 1\} \quad \forall s \in SC, t \in T, m \in M, c \in C \]  
(7.16)

\[ \rho_{s, t}^{m, c} \in \{0, 1\} \quad \forall s \in SC, t \in T, m \in M, c \in C \]  
(7.17)

\[ y_{l, s, t}^{m, c} \in Z^+ \quad \forall s \in SC, t \in T, m \in M, c \in C \]  
(7.18)

\[ h_{a, \tau}^{3-4} \in R^+ \quad \forall a \in A, \tau \in TD_a \]  
(7.19)

\[ h_{a, \tau}^{4-up} \in R^+ \quad \forall a \in A, \tau \in TD_a \]  
(7.20)

\[ d_{a, \tau} \in R^+ \quad \forall a \in A, \tau \in TD_a \]  
(7.21)

\[ y_{a, m} \in Z^+ \quad \forall m \in M \]  
(7.22)

Constraints (7.13)-(7.22) state the variables’ dominion.

7.5 Computational Experiments

All of our computational experience is for realistic cases drawn from RENFE’s regional network in Madrid, also known as ”Cercanías Madrid” (Figure 4.1). The network is characterized by its modular structure. That is, in real life it is separated into different and independent modules for operating purposes. The capacity assigned to each module is decided jointly with the infrastructure manager during the RT phase. That is, the operator designs a planning with given capacities. This planning is presented to the infrastructure manager and the planning may be accepted or rejected. If it is rejected the timetable must be updated in order to match the infrastructure manager’s willing. Thus, the operator faces an iterative process to develop the RT and therefore the RS assignment.

The model presented in this chapter has been tested in lines C3 (Figure 5.2), C4 (Figure 5.3) and C5 (Figure 5.1). The available rolling stock’s characteristics are showed in Tables 5.1 and 5.6.

Inside line C5 we consider the possibility of attending 10 different train lines. These train lines are defined by their initial and final depot stations and every station they come through. They are shown in Table 7.1. In the first column the line number is shown, in the second one the departure depot station and in the last one the arrival depot station.
Table 7.1 Train lines in Line C5

<table>
<thead>
<tr>
<th>Lines</th>
<th>Origin</th>
<th>Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>ℓ1</td>
<td>Humanes</td>
<td>Atocha</td>
</tr>
<tr>
<td>ℓ2</td>
<td>Atocha</td>
<td>Humanes</td>
</tr>
<tr>
<td>ℓ3</td>
<td>Fuenlabrada</td>
<td>Atocha</td>
</tr>
<tr>
<td>ℓ4</td>
<td>Atocha</td>
<td>Fuenlabrada</td>
</tr>
<tr>
<td>ℓ5</td>
<td>Humanes</td>
<td>Móstoles El Soto</td>
</tr>
<tr>
<td>ℓ6</td>
<td>Móstoles El Soto</td>
<td>Humanes</td>
</tr>
<tr>
<td>ℓ7</td>
<td>Fuenlabrada</td>
<td>Móstoles El Soto</td>
</tr>
<tr>
<td>ℓ8</td>
<td>Móstoles El Soto</td>
<td>Fuenlabrada</td>
</tr>
<tr>
<td>ℓ9</td>
<td>Atocha</td>
<td>Móstoles El Soto</td>
</tr>
<tr>
<td>ℓ10</td>
<td>Móstoles El Soto</td>
<td>Atocha</td>
</tr>
</tbody>
</table>

Table 7.2 Train lines in Lines C3&C4

<table>
<thead>
<tr>
<th>Lines</th>
<th>Origin</th>
<th>Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>ℓ1</td>
<td>Parla</td>
<td>Colmenar</td>
</tr>
<tr>
<td>ℓ2</td>
<td>Colmenar</td>
<td>Parla</td>
</tr>
<tr>
<td>ℓ3</td>
<td>Parla</td>
<td>Alcobendas</td>
</tr>
<tr>
<td>ℓ4</td>
<td>Alcobendas</td>
<td>Parla</td>
</tr>
<tr>
<td>ℓ5</td>
<td>Parla</td>
<td>Chamartín</td>
</tr>
<tr>
<td>ℓ6</td>
<td>Chamartín</td>
<td>Parla</td>
</tr>
<tr>
<td>ℓ7</td>
<td>Colmenar</td>
<td>Atocha</td>
</tr>
<tr>
<td>ℓ8</td>
<td>Atocha</td>
<td>Colmenar</td>
</tr>
<tr>
<td>ℓ9</td>
<td>Alcobendas</td>
<td>Atocha</td>
</tr>
<tr>
<td>ℓ10</td>
<td>Atocha</td>
<td>Alcobendas</td>
</tr>
<tr>
<td>ℓ11</td>
<td>Aranjuez</td>
<td>Chamartín</td>
</tr>
<tr>
<td>ℓ12</td>
<td>Chamartín</td>
<td>Aranjuez</td>
</tr>
</tbody>
</table>

Inside lines C3 and C4 we consider the possibility of attending different train lines. These train lines are defined by their initial and final depot stations and every station they come through. They are shown in Table 7.2. In the first column the line number is shown, in the second one the departure depot station and in the last one the arrival depot station.

In a daily planning period from 5:00 a.m. to 1:00 a.m. divided into one minute periods, we have 1200 time periods. The rotation time is 3 minutes at every depot station.

Our runs were performed on a Personal Computer with an Intel Core2 Quad Q9550 CPU at 2.83 GHz and 8 GB of RAM, running under Windows 7 64Bit, and our programs were implemented in GAMS/Cplex 11.1.
Table 7.3 Line C5 RT&RS model: the number of variables, constraints and non-zeros

<table>
<thead>
<tr>
<th></th>
<th>RT &amp;RSM</th>
</tr>
</thead>
<tbody>
<tr>
<td># Discrete variables</td>
<td>54130</td>
</tr>
<tr>
<td># Continuous variables</td>
<td>190804-190384-189409-188113-186847</td>
</tr>
<tr>
<td># Constraints</td>
<td>105386-104826-103526-101798-100110</td>
</tr>
<tr>
<td># Non-zero elements</td>
<td>1396843-1601563-1758451-1853187-1902275</td>
</tr>
</tbody>
</table>

7.5.1 Passenger Demand Approach

Passenger demand is defined as a flow through every arc $a$ during each time period $\tau$. In $\tau$’s definition the scalar $k$ is implicit. This scalar’s value represents the level of aggregation for passenger demand. For example, for $k = 1$ we have the union of two consecutive time periods $t, t'$; that is, we will have that every passenger flow departing during $t, t'$ will be treated as an aggregated passenger flow.

In this way, depending on $k$’s value, the model size and solution will change. In order to illustrate the RT&RS model size for different $k$ values, Table 7.3 shows the model size for the Line C5 study case. The RT&RSM numbers of discrete and continuous variables, constraints and non-zero elements are given for the model (RT&RSM). The number of discrete variables remains equal for every case because they do not depend on $\tau$. However, the rest of variables, constraints and non-zero elements vary with $\tau$ definition.

We have done a study varying $\tau$ definition. The different cases we have studied are the following: aggregating two time periods ($k = 1$), four ($k = 3$), six ($k = 5$), eight ($k = 7$) and ten ($k = 9$) time periods. In this way, the model size is presented for every studied case in Table 7.3 for Line C5.

Every computational experience has been stopped after 1 hour of computational time. For every of them the optimality gap was always lower than 3%. In these computational experiments the robustness parameters have not been included. That is, the parameter $\theta_{s,s',t}$ is set to one and the parameter $\vartheta_{s,t}$ to its nominal value. On the left handside of Figure 7.1 the objective function value of the mixed-integer program and of the linear relaxation program is showed for Line C5. The first one is represented by the continuous line and the second one by the discontinuous line. The value corresponding to $k = 0$ is the case where the RS assignment is done for a given and fixed timetable. As $k$ goes up we can see how the liner relaxation becomes weaker; we must mention that the optimality gap (after 1 hour of computation) also increases as $k$ goes up, but never exceeding the 3%. TSOC are represented by the dotted line, which maintains the same behaviour as the objective function for the mixed-integer program. On the right handside,
the number of convoys used (continuous line) and the number of composition changes performed (discontinuous line) are shown for every case. The convoys used is maintained almost equal for every case. However, the number of performed composition changes is strongly reduced as \( k \) goes up. One could think that decreasing composition changes mandates a bigger fleet. Nevertheless, composition changes decrease as \( k \) increases, that is, as the time window’s span we are using to attend the passenger demand increases, which means that the model has more freedom to adequate RS flows without performing composition changes.

![Figure 7.1](image)

**Figure 7.1 Different RT&RS solutions for Line C5 obtained by varying \( \tau(k) \) value**

The minimum separation time imposed between two consecutive trains using the same infrastructure is given by the network operator and the infrastructure manager. For Line C5 this value is of 3 minutes equivalent to 3 time periods. The value of this parameter is directly related with the choice of \( k \). We must ensure that trains are coming through the different arcs and time periods \( \tau \) enough separated in order to reduce passengers in excess. That is, if we choose \( \tau \)’s length lower than the time separation, it may occur that many passengers are neglected. In Figure 7.2 passengers in excess and neglected passengers for the two consecutive arcs with the greatest passenger flow are showed for Line C5. The frequency values for these arcs are also represented. The continuous line represents these frequency values which are maintained around the same value for every tested case. The discontinuous and dotted lines represent the sum of neglected and in excess passengers for two different arcs. We must mention that neglected passengers appeared only for the case where \( k = 1 \). In order to develop more computational experiments for the Line C5 we choose the case where \( \tau \) is exactly two times the value of the separation time, that is, \( k = 5 \). In this case we can see how passengers in excess are halved compared to the initial case where the timetable was fixed.

For Lines C3&C4 the same approach has been used, that is, 1 hour of computational time for each study case. For every of them the optimality gap was also always lower than 3%. On the left handside of Figure 7.3 the objective function value of the mixed-
Figure 7.2 Passengers in excess for the arcs with the greatest passenger flow in Line C5 integer program and of the linear relaxation program is showed for Lines C3&C4. The first one is represented by the continuous line and the second one by the discontinuous line. The value corresponding to \( k = 0 \) is the case where the RS assignment is done for a given and fixed timetable. TSOC are represented by the dotted line, which maintains the same behaviour as the objective function for the mixed-integer program. On the right handside, the number of convoys used (continuous line) and the number of composition changes performed (discontinuous line) are shown for every case.

Figure 7.3 Different RT&RS solutions for Lines C3&C4 obtained by varying \( \tau (k) \) value

In Lines C3&C4 the separation time is of 5 minutes. As we said before, if we choose \( \tau \)'s length lower than the time separation, it may occur that many passengers are neglected. For this reason the objective function value in Figure 7.3 is very high for \( k \) ranging from 1 to 3. In Line C5 this issue was not a big problem because the separation time was lower and even for \( k = 1 \) there were few neglected passengers. In order to develop more computational experiments for Lines C3&C4 we choose the case where \( \tau \) is exactly two times the value of the separation time, that is, \( k = 9 \).

In Table 7.4 different solutions are listed for Lines C5 and C3&C4. For each of them, in the first row the solution obtained by the proposed RT&RS model is shown (for Line C5 \( k = 5 \) and for Lines C3&C4 \( k = 9 \)). In the second one, the solution obtained with the RS assignment with fixed timetable. In the last one, the current solution operated by RENFE. In the third column the number of used train units (#C) is listed; this number
Table 7.4 RT&RS model solution compared to RS with fixed timetable solution and current RENFE solution

<table>
<thead>
<tr>
<th>Line</th>
<th>Solution</th>
<th>#C</th>
<th>TSOC</th>
<th>EMOC</th>
<th>PEC</th>
<th>#CC</th>
<th>Savings (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5 (k = 5)</td>
<td>RT&amp;RS</td>
<td>57</td>
<td>75792.24</td>
<td>1892.72</td>
<td>2464</td>
<td>8</td>
<td>30.7</td>
</tr>
<tr>
<td></td>
<td>RS</td>
<td>64</td>
<td>80099.76</td>
<td>1265.04</td>
<td>3554</td>
<td>20</td>
<td>27.4</td>
</tr>
<tr>
<td></td>
<td>Current</td>
<td>74</td>
<td>109765.20</td>
<td>2232.12</td>
<td>874</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>C3&amp;C4 (k = 9)</td>
<td>RT&amp;RS</td>
<td>48</td>
<td>81865.91</td>
<td>2086.29</td>
<td>1372</td>
<td>12</td>
<td>41.2</td>
</tr>
<tr>
<td></td>
<td>RS</td>
<td>63</td>
<td>87837.52</td>
<td>4090.83</td>
<td>2177</td>
<td>34</td>
<td>35.6</td>
</tr>
<tr>
<td></td>
<td>Current</td>
<td>72</td>
<td>136633.86</td>
<td>6083.88</td>
<td>1550</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

is lowered compared to the current solution operated by RENFE and to the one obtained by RS assignment with fixed timetable. In this way, maintenance costs are also reduced because they depend on the number of used convoys. In the fourth column operating costs (TSOC) are shown, which are also slightly reduced. In the fifth column, empty movement costs (EMOC) are listed; they are reduced compared to the current solution; however, in the RT&RS solution these costs are increased compared to the RS assignment with fixed timetable. Passengers in excess costs are increased compared to the current solution; however, these costs include passengers above 3 $\text{pax/m}^2$ and RENFE considers passengers in excess when there are passengers exceeding 3.5 $\text{pax/m}^2$ (see Cadarso and Marín [34]). In the current solution there are no composition changes (#CC); however, in order to enable a schedule using fewer convoys they arise for the studied cases; for the case where the timetable is optimized the number of composition changes is lower due to the great freedom that supposes the timetable optimization. In the last column savings in operating costs with respect to the current solution are shown.

7.5.2 Maintenance Costs

An important cost to the operator is the maintenance cost. For example, for this material, there is a daily fixed cost of, say, 400 €. Given the importance of the number of train units used in the network, different solutions must account for that number.

If we compare any of the proposed solutions in Table 7.4 with the current solution provided by RENFE’s operators, we can see that the number of used train units is smaller.

However, the model does not account for this maintenance costs. These costs may be included in the model formulation through an economic penalty representing maintenance costs per used convoy in the objective function. This cost is represented by a parameter $mc_m$ which depends on the material type. The complete term in the objective function
representing daily maintenance costs is as follows: $\sum_{s,m,c \in S,M,C} mc_{m}o_{c}yt_{s,t,i,m,c}$. It is the sum of the number of train units of each material in every station at the beginning of the planning period multiplied by the daily maintenance cost.

In Table 7.5 the obtained results are presented. For each study case, the RT&RSM with maintenance costs solution is listed first. Secondly the RT&RSM solution. As maintenance costs depending on the number of train units used have been included, the number of train units used is lower than in the case where they are not. However, TSOC are increased; this may be due to the fact that as fewer train units are available, capacity cannot be adjusted as in the case where more convoys were available, so more kilometers must be rolled by the fleet. PEC present an expected behavior because as capacity is not so well adjusted they go up. Finally, in order to allow using a smaller fleet size more composition changes are performed.

### 7.5.3 Robust Solutions

Robustness parameters are obtained from operators. For example, in line C5, the Atocha depot station is shared among more than 5 different lines. This causes the capacity of C5 material to vary strongly during the planning period. Robustness parameters are chosen to try to avoid (if possible) composition changes and empty movements with the destination Atocha.

As we have stated above, robustness may be introduced with different approaches. Some of these approaches could prevent dangerous empty movements and composition changes. Dangerous empty movements and composition changes are recognized as those occurring during rush hours. To illustrate this effect, two different cases are shown. In the first case, no robustness is introduced (NoRob) (i.e., $\theta_{s,s',t} = 1$ and $\vartheta_{s,t}$ is equal to its nominal value for all possible cases). In the second case, robustness is introduced (Rob), doubling the values of these parameters at rush hours from 7:00 a.m. to 10:00 a.m. and from 2:00 p.m. to 5:00 p.m. The computational results are shown in Table 7.6. In both cases, the rest of the parameters are those used in the previous examples.
We can see the differences between both cases in Table 7.6. In the second column, the train service operating costs (TSOC) are shown. In the robust case, these costs are greater because more train services are set to double composition; this cost may represent the robustness cost. In the third column, the empty movement costs (EMOC) are shown, which are lower than those of the no robustness case. If we pay attention to empty movements during rush hours (#EMRH), we can see that the number of empty movements is reduced, achieving one of our goals. In a similar way, robustness is introduced in the solution by avoiding composition changes during rush hours (#CCRH). However, introducing robustness into the model makes passengers in excess costs (PEC) to rise up. In this case, robustness costs remain partially on passengers.

Another way of introducing robustness into the model is to ensure swapping operations. The model is obeyed to provide solutions with free parkings in depot stations at lines’ edges. In this way, the obtained solution will be more resistant to possible delays because they may be alleviated by swapping trains. This robustness (S-Rob) may be achieved by reducing strategic depot stations capacity in order to always allocate some convoys to perform swapping operations.

We compare robust (Rob) and swapping robust (S-Rob) solutions in Table 7.7. Robust solutions are those ones obtained applying the previous explained approach. Swapping robust solutions are those ones obtained by ensuring capacity in certain depot stations in order to perform swapping operations if necessary. As we can see, TSOC are showed in the second column. EMOC in the third one. #EMRH, in the fourth column, are strongly reduced for Line C5 but they are strongly increased for Lines C3&C4. These two opposite solutions mean that in Line C5 there is enough capacity to continue operating even when we reduce capacity in order to ensure swapping operations. However, for Lines C3&C4 this is not true, capacity is not enough and many empty movements must be done during rush hours. In this way, we are gaining robustness through swapping opportunities but we are also getting worse robustness through empty movements: a trade off must be selected depending on the available infrastructure. In the next column, PEC are showed. They are reduced due to the fact that more trains are set to double composition, and that more
Table 7.7 Comparing robust and swapping robustness solutions for line C5

<table>
<thead>
<tr>
<th>Case</th>
<th>TSOC</th>
<th>EMOC</th>
<th>#EMRH</th>
<th>PEC</th>
<th>#CC</th>
<th>#CCRH</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5 Rob</td>
<td>76966.72</td>
<td>650.64</td>
<td>10</td>
<td>3311</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>C5 S-Rob</td>
<td>77302.08</td>
<td>646.24</td>
<td>3</td>
<td>2728</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>C3&amp;C4 Rob</td>
<td>82847.21</td>
<td>1816.62</td>
<td>1</td>
<td>1658</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>C3&amp;C4 S-Rob</td>
<td>82697.69</td>
<td>2884.74</td>
<td>11</td>
<td>1225</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

Composition changes are needed; both facts make to fit better capacity to passenger flows. Composition changes (#CC) are increased because capacity for the S-Rob case is lower. Similarly, composition changes during rush hours (#CCRH) are increased.

7.6 Summary

The integration of the railway timetabling and rolling stock assignment problems is presented in this chapter. The planned frequencies are known from the railway line planning problem. In the model operating costs are minimized jointly with penalties for robustness. Robustness has been defined through penalizing difficult shunting operations. Another robust approach has also been included: ensuring parking in strategic depot stations in order to perform swapping operations. Maintenance costs have been included adding to the objective function maintenance daily costs for each train unit.

Computational results show how the rolling stock assignment depends on the efficiency of the railway timetable. We do not only change the timetable; also updating frequency values we introduce a greater freedom. That is, comparing with the rolling stock assignment with fixed timetable, we can see how better solutions can be obtained improving the overall solution efficiency and getting a greater robustness degree.
Chapter 8

IMPROVING ROBUST ROLLING STOCK CIRCULATION IN RAPID TRANSIT NETWORKS

The rolling stock circulation depends on two different problems: the rolling stock assignment and the train routing problems, which up to now have been solved sequentially. Therefore, the integration of these decision making is justified and is appropriate to introduce robustness in the model. We propose a new approach to obtain better and more robust circulations of the rolling stock train units, solving the rolling stock assignment accounting for the train routing problem. However, the integrated approach provides a huge model. Then, we solve the integrated model using Benders decomposition, where the main decision is the rolling stock assignment and the train routing is in the second level. For computational reasons we propose a heuristic based on Benders decomposition. Computational experiments show how the current solution operated by RENFE can be improved: more robust and efficient solutions are obtained.

8.1 Introduction

Rolling Stock (RS) circulations are completely determined once the RS assignment and Train Routing (TR) problems are solved. This chapter presents a mathematical model for improving the Robust Circulation of the Rolling Stock (RCRS) in rapid transit networks. The word circulation refers to both the RS assignment and the TR problems. Up to now, RS circulations have been determined in a sequential manner. First, the RS assignment was found and then, the routing of the train units was developed. However, this sequential
approach may lead to sub-optimal solutions because the RS assignment and shunting operations were planned without accounting for delays. Consequently, if a train unit arrives late to perform the next service a swapping operation will be needed; and if this is not possible, the operator will face a disruption. In the integrated approach the RS assignment and shunting operations are designed accounting for delays; this means that the number of swapping operations will be minimized obtaining robust and improved RS circulations. Robustness aspects may be also introduced with different criteria. From the RS assignment point of view difficult shunting operations may be avoided. From the TR problem point of view human resources may be optimized. If excessive extra human resources are needed for train units sequences they will not be available in case of disruption to drive any other train.

This chapter is organized as follows. A literature overview is given in Section 8.2. We describe the rolling stock circulation problem in Section 8.3. In Section 8.4 the mathematical formulation is presented in detail. Section 8.5 contains the solution approach based on Benders Decomposition. Section 8.6 shows computational results based on a realistic case provided by RENFE.

8.2 Literature Review

An integer programming model is considered by Alfieri et al. [3] to determine the RS circulation for multiple RS types on a single line and on a single day. They use the concept of a transition graph to deal with this aspect. The objective is to minimize the number of units or the carriage-kilometers such that the given passenger demand is satisfied. The approach is tested on real-life examples from NS, the main operator of passenger trains in the Netherlands. Maróti [94] focuses on planning problems that arise at NS. He identifies tactical, operational and short-term rolling stock planning problems and developed operations research models for describing them. Cadarso and Marín [34] propose a mixed integer optimization model to study suburban rapid transit robust RS assignment. They minimize total costs where the costs include service trips, robustness-relevant empty train movements and composition change costs.

De Almeida et al. [57] study the concept of robustness in railway production planning. They state that robustness can be improved by reducing the propagation of delays and increasing the number of feasible resource allocation exchanges. Cadarso and Marín [35] present an integer programming model to determine a sequence of operations to be rolled by the train units such that each operation is included exactly in one sequence and there
is always the number of necessary train units available for every operation execution.

Cacchiani et al. [42] describe a two-stage optimization model for determining robust rolling stock circulations for passenger trains. Here robustness means that the rolling stock circulations can better deal with large disruptions of the railway system. The two-stage optimization model is formulated as a large mixed-integer linear programming model. They evaluate their approach on the real-life rolling stock-planning problem of NS, the main operator of passenger trains in the Netherlands.

Cadarso and Marín [36] propose an integrated planning model to adequate the offered capacity and system frequencies to attend the increased passenger demand and traffic congestion around urban and suburban areas. The railway timetable and rolling stock assignment are integrated. They ensure swapping operations using homogeneous rolling stock and ensuring parkings in strategic stations. They illustrate the model using computational experiments drawn from RENFE in Madrid, Spain.

In the airline industry, some authors have addressed the integrated fleet assignment and routing problem. A daily scheduling and routing problem is addressed by Desaulniers et al. [58]. They propose two equivalent formulations: a nonlinear multicommodity flow formulation and a set partitioning formulation. Barnhart et al. [6] present a string-based model and a branch-and-price solution approach that accommodates the costs associated with aircraft connections as well as complicating maintenance constraints.

8.3 Rolling Stock Circulation Problem Description

The RS circulation is completely determined with the RS assignment and TR problems. 

**Rolling Stock Assignment.** The goal of the RS assignment problem is to determine train units’ type and number of them that compose trains considering a given timetable and a demand to satisfy in a context in which shunting is optimized. Cadarso and Marín [34] considered the railway rolling stock assignment problem for rapid transit networks.

**Train Routing.** Train routing is the process of assigning each individual train unit, referred to as an identification number, to RS operations. Given the RS assignment, we must determine a sequence of operations to be rolled by an individual train unit such that the assigned operations are included in exactly one sequence. Cadarso and Marín [33] considered the railway rolling stock routing problem for rapid transit networks.

**Robustness.** Robustness is introduced penalizing composition changes and empty services. The system is made more robust by assigning only one material type per line (i.e., for every train service operating the same line, the material must be equal). This con-
straint allows for all material on one line to be swapped between different train services at depot stations serving that line. Thus, there will be more opportunities to swap train services if an incident occurs, and the propagation of the incident can be mitigated in an easier way.

The propagated delay from one operation to another operation (operation connection) is defined by $pd = \max(ad - slack, 0)$, where $ad$ is an aleatory variable representing the arrival delay (see Cadarso and Marín [33]), and $slack$ is the planned slack between both operations. When $pd > 0$, the train unit performing the first operation will not be on time to perform the following operation, that is, the operation connection cannot be performed. Thus, this robustness criterion indirectly minimizes the number of necessary swapping operations and minimizes the human resources required to perform the material swapping.

Another issue affecting train units performance is their ability to be ready to perform a service (see Cadarso and Marín [33]). Human resources are needed to prepare train units if they have been stopped for a while. Thus, robustness may be also introduced through penalizing crew requirement. With this criterion more human resources will be available in case of disruption in the network because their requirement is minimized when performing sequences.

8.4 Robust Circulation of the Rolling Stock Model

Railway planning is currently divided into several optimization steps from first strategic decisions to daily operations. It is well known that disintegrated planning produce optimal solutions for each stage but non-optimal global solutions. Up to day, rolling stock assignment and train routing (rolling stock circulations) have been separately considered in rapid transit networks. A new integrated model is proposed for these two phases in order to obtain better circulations. Moreover, a greater robustness degree is obtained through this integrated approach: consider two consecutive planning stages. In order to obtain a high-quality solution in the second stage, some slack must be introduced in the first stage. In the same way, a robust solution adds slack to safeguard against data perturbation. This slack may provide a smooth interface between subsequent planning stages. This fact becomes a connection between robustness and integration.

The Robust Circulation of the Rolling Stock Model (RCRSM) described below is formulated as a multicommodity flow model. Before introducing the mathematical formulation, the following notation is explained:
- Sets:
  
  - $L$: set of services. Each service is characterized by an origin, a destination and a departure time.
  
  - $L^t \subset L$: subset of train services. These services are performed to attend the passenger demand.
  
  - $L^e \subset L$: subset of empty services. These services cannot attend passenger demand.
  
  - $S$: set of nodes. The nodes are defined by a station and a time period.
  
  - $SC \subset S$: subset of depot nodes.
  
  - $CS \subset S$: subset of nodes at which the material is counted.
  
  - $A$: set of arcs. They are characterized by a departure station and time period and an arrival station.
  
  - $M$: set of train units types.
  
  - $C$: set of compositions.
  
  - $C_m \subset C$: subset of compositions of material $m$.
  
  - $A_\ell$: set of arcs $a$ attended by train service $\ell \in L^t$.
  
  - $I$: set of operations. There are three types of operations: services (indexed by 1), aggregations (indexed by 2) and disaggregations (indexed by 3). We mean by services train and empty services. The last two types of operations refer to composition changes. This set is indexed by $i$ and $j$.

- Parameters:

  - $oc_c$: operating cost for composition $c$ per rolled kilometer.
  
  - $km_\ell$: kilometers in service $\ell$.
  
  - $dpc_{a,\ell}$: disrupted passenger cost in arc $a$ used by train service $\ell \in L^t$.
  
  - $\vartheta_s$: penalty cost for composition change in node $s$.
  
  - $\psi, \zeta$: cost per train delay time period and per human necessity and time period, respectively.
• $E[pd_{i,s}^{j,s'}]:$ expected delay time periods propagated from operation $i$ ending at node $s$ to operation $j$ beginning at node $s'$.

• $cr_{i,s}^{j,s'}$: penalty for extra human resources necessity to enable train units to be ready between operation $i$ ending at node $s$ and operation $j$ beginning at node $s'$.

• $pf_{a,\ell}$: expected passenger flow in arc $a$ used by train service $\ell \in L^i$.

• $cap_c$: passenger capacity in composition $c$.

• $\alpha_{\ell,s}$: $=-1$, if service $\ell$ leaves from node $s$; $=1$, if service $\ell$ arrives at node $s$; $=0$, otherwise.

• $\chi_m$: fleet size for convoys of type $m$.

• $e, d$: time needed for aggregation and disaggregation, respectively.

• $tu_c$: number of train units in composition $c$.

• $\beta_{\ell}$: $=1$, if service $\ell$ is rolling at the count time period; $0$, otherwise.

• $cn_{c,c'}$: number of compositions $c$ needed to obtain a composition $c'$ in case of aggregation (number of compositions $c'$ obtained from composition $c$ in case of disaggregation).

• $\mu_{s',s}$: $=1$, if a composition change which started in node $s'$ is still being performed in node $s$.

• $s_i, s_f$: initial and final nodes in the planning period.

• $\tilde{\alpha}_{i,c}$: $=-2$, if $i$ is a disaggregation (index 3) followed by trains with $c$ composition; $=-1$, otherwise.

• $\tilde{\beta}_{j,c}$: $=-2$, if $j$ is an aggregation operation (index 2) preceded by trains with $c$ composition; $=1$, otherwise.

- Variables:

• $x_{c,\ell}$: $=1$, if train service $\ell \in L$ uses composition $c$; $=0$, otherwise.

• $y_{c,s}$: integer variable. Train inventory of composition $c$ in node $s$.

• $dpc_{a,\ell}$: positive variable. Disrupted passengers in arc $a$ used by train service $\ell \in L^i$. 
• $\epsilon_s^{c,c'}$, $\left(\delta_s^{c,c'}\right)$: $=1$, if we begin the aggregation (disaggregation) at node $s$, from composition $c$ to composition $c'$; $=0$, otherwise.

• $cc_s^{c,c'}$: $=1$, if we begin a composition change at node $s$, from composition $c$ to composition $c'$; $=0$, otherwise.

• $seq_{i,s}^{j,s',c}$: $=1$, if operation $i$ ending at node $s$ is followed by operation $j$ beginning at node $s'$, both of them with composition $c$; $=0$, otherwise.

• $\varphi_{i,s}^{c}$, $\left(\varphi_{j,s}^{c}\right)$: positive variable. It determines the number of operations $i$ that end ($j$ that begin) in node $s$ with composition $c$; $=0$, otherwise.

The RCRSM mathematical formulation is as follows:

**8.4.1 Objective Function**

\[
\min z = \sum_{\ell \in L} \sum_{c \in C} oc_{\ell} km_{\ell} x_{s}^{c} + \sum_{\ell \in L} \sum_{a \in A_{\ell}} dpc_{a,\ell} dp_{a,\ell} + \sum_{s \in SC} \sum_{c,c' \in C} \vartheta_{s} \cdot cc_{s}^{c,c'} + \\
\sum_{i,j \in I} \sum_{s,s' \in S} \sum_{c \in C} \left[ \psi \cdot E \left[ pd_{i,s}^{j,s'} \right] + \zeta \cdot cr_{i,s}^{j,s'} \right] \cdot seq_{i,s}^{j,s',c} \tag{8.1}
\]

In the objective function, different costs are minimized:

1. services operating costs; for dangerous empty services $\ell \in L^e$ the value $km_{\ell}$ is increased in order to introduce robustness;

2. disrupted passengers costs, that is, every passenger who cannot board the train is assumed to be a cost for the operator;

3. shunting costs; the parameter $\vartheta_{s}$ is modulated depending on the shunting operation, that is, if a shunting operation is likely to fail, the associated cost to it will be increased to introduce robustness;

4. expected delay; for calculating the $pd_{i,s}^{j,s'}$ values, expressed in minutes, we take advantage of the fact that the slack between operations is always a constant value. Therefore, this value can be captured through the location parameter $\theta_{i,s}^{j,s'}$, and the propagated delay is calculated in (8.2) (see Cadarso and Marín [33]).

5. and extra human resources necessity to enable train units to be ready between operations; the penalty $cr_{i,s}^{j,s'}$ varies with the time that the train unit has been stopped.
First, it grows linearly with time, and after a while it is constant. Consequently, a piecewise penalization function is used for $cr_{i,s}^{j,s'}$.

$$E[pd_{i,s}^{j,s'}] = \theta_{i,s}^{j,s'} \cdot \left[ 1 - \phi \left( \frac{\ln \left( \frac{a - \theta_{i,s}^{j,s'}}{m_{i,s}} \right)}{\sigma_{i,s}} \right) \right] + m_{i,s} \cdot e^{-\frac{\sigma_{i,s}^2}{2}} \cdot \left[ 1 - \phi \left( \frac{\ln \left( \frac{a - \theta_{i,s}^{j,s'}}{m_{i,s}} \right)}{\sigma_{i,s}} \right) - \sigma_{i,s} \right]$$  \hspace{1cm} (8.2)

where $\phi(x)$ is the cumulative distribution function of a standard normal distribution, $m_{i,s}$ is the scale factor, $\sigma_{i,s}$ is the standard deviation, and $a$ is equal to 0 if $\theta_{i,s}^{j,s'} \leq 0$ or $a$ is equal to $\theta_{i,s}^{j,s'}$, otherwise.

Decision variables are subject to the constraints as follows.

### 8.4.2 Passengers Constraints

The following group of constraints links the allocated seat capacity to the number of passengers $pf_{a,\ell}$.

$$\sum_{c \in C} cap_c x_{\ell}^c \geq pf_{a,\ell} - dp_{a,\ell} \quad \forall \ell \in L^l, a \in A_{\ell}$$  \hspace{1cm} (8.3)

The constraints (8.3) say that for each arc $a \in A$ attended by train service $\ell \in L^l$, the capacity of the train is enough to accommodate the passenger demand minus the denied passengers.

### 8.4.3 Rolling Stock Constraints

$$\sum_{c \in C} x_{\ell}^c = 1 \quad \forall \ell \in L^l$$  \hspace{1cm} (8.4)

$$\sum_{c \in C} x_{\ell}^c \leq 1 \quad \forall \ell \in L^e$$  \hspace{1cm} (8.5)

Constraints (8.4) state that each train service $\ell \in L^l$ must be assigned a composition $c$. Constraints (8.5) express that empty services $\ell \in L^e$ get at most one composition.
Inventory conservation constraints (8.6) ensure the train units’ flow balance. These constraints consider the increases or decreases of the inventory depending on departing or arriving services and also on local shunting operations. Obviously, the inventory is always nonnegative.

Fleet capacity constraints (8.7) ensure that the number of train units used at the count time period is limited by the size of the fleet. Note that these constraints count the running trains and those ones in depot stations. Depot capacity constraints (8.8) ensure that the total capacity of the station is not exceeded.

Constraints (8.9) count the number of composition changes in every depot station. Note that for all the composition changes that are not physically possible (i.e., due to composition incompatibility), the variables $\epsilon_s^{c'}$, $\delta_{s}^{c,c'}$ are fixed to zero value. Constraints (8.10) denote that the inventory during the initial and final nodes must be equal. Consequently, the obtained schedule is periodic and may be repeated at the next day.
8.4.4 Coupling Constraints

\[ x_{c}^\ell = \varphi_{1,s} \quad \forall \ell \in L, c \in C, s \in SC : \alpha_{\ell,s} = -1 \] (8.11)
\[ x_{c}^\ell = \phi_{1,s} \quad \forall \ell \in L, c \in C, s \in SC : \alpha_{\ell,s} = 1 \] (8.12)
\[ \epsilon^{c,c'}_s = \varphi^{2,s}_c \quad \forall s \in SC, c, c' \in M, C \] (8.13)
\[ \epsilon^{c,c'}_s = \phi^{2,s+e}_{2,s} \quad \forall s \in SC, c, c' \in C \] (8.14)
\[ \delta^{c,c'}_s = \varphi^{3,s}_c \quad \forall s \in SC, c, c' \in C \] (8.15)
\[ \delta^{c,c'}_s = \phi^{3,s+d}_{3,s} \quad \forall s \in SC, c, c' \in C \] (8.16)

Constraints (8.11)-(8.16) determine the characteristics of each operation: the node at which it starts and ends and the composition assigned. This information is stored in the variables \( \phi_{1,s}, \varphi_{1,s} \). For example, in constraints (8.11) each time that variable \( x_{c}^\ell \) takes value 1, the right hand side of the constraint will also be one. Therefore, the variable \( \varphi_{1,s} \) will take value 1: this means that a service starts at \( s \) with composition \( c \).

8.4.5 Sequence Constraints

\[ \sum_{i \in I} \sum_{s' \in S} \text{seq}_{i,s'}^{j,s} = \tilde{\alpha}_{i,c} \phi_{1,s} \quad (\kappa_{i,s}^{c}) \quad \forall i \in I, s \in S, c \in C \] (8.17)
\[ \sum_{i \in I} \sum_{s' \in S} \text{seq}_{i,s'}^{j,s} = \tilde{\beta}_{j,c} \varphi_{1,s} \quad (\lambda_{j,s}^{c}) \quad \forall j \in I, s \in S, c \in C \] (8.18)

Constraints (8.17)-(8.18) are sequencing constraints. They ensure that every operation is preceded by another one and that every operation is continued by other operation. Here, the variables \( \tilde{\alpha}_{i,c}, \tilde{\beta}_{j,c} \) are used. The former introduces \( i \)'s operation ending information in order to find a following compatible operation. Similarly, the latter shows beginning information of operation \( j \) to find a preceding compatible operation. \( \kappa_{i,s}^{c}, \lambda_{j,s}^{c} \) are dual variables.

The RCRSM presented formulation leads us to a huge model size. For a real instance in the regional network in Madrid, we have with this new integrated formulation more than one hundred million binary variables. Therefore, we decompose it with Benders Decomposition.
8.5 Solution Approach: Benders Decomposition

Benders decomposition may be obtained by classifying variables in difficult and easy variables. Difficult variables are RS variables \([x_{c,\ell}, y_{c,s}, dpc_{a,\ell}, \epsilon_{c,\ell}, \delta_{c,\ell}, \epsilon_{c,c'}, \phi_{i,s}, \varphi_{j,s}]\). That is, once difficult variables are known it is relatively easy to determine the optimal sequence associated to them. So, easy variables are sequence variables \(seq_{i,s}^{j,s',c}\). Hence, we have that the master model \((MM)\) will be the passengers constraints, RS constraints, coupling constraints and Benders optimality cuts; and the submodel \((SM)\) the Train Routing \((TR)\) model.

8.5.1 Benders Submodel

The Benders \(SM\) at each iteration \(it\) \((SM^{it})\) will be the train routing model:

\[
\begin{align*}
\min z &= \sum_{i,j \in I} \sum_{s,s' \in S} \sum_{c \in C} \left[ \psi \cdot E \left[ pd_{i,s}^{j,s'} \right] + \zeta \cdot cr_{i,s}^{j,s'} \right] \cdot seq_{i,s}^{j,s',c} \\
\sum_{j \in I} \sum_{s' \in S} seq_{i,s}^{j,s',c} &= \tilde{\alpha}_{i,c} \phi_{i,s}^{it} \quad (\kappa_{i,s}) \forall i \in I, s \in S, c \in C \\
\sum_{i \in I} \sum_{s' \in S} seq_{i,s}^{j,s',c} &= \tilde{\beta}_{j,c} \varphi_{j,s}^{it} \quad (\lambda_{j,s}) \forall j \in I, s \in S, c \in C
\end{align*}
\]

where \(\phi_{i,s}^{it}, \varphi_{j,s}^{it}\) is the \(MM\) solution at iteration \(it\) \((MM^{it})\). As we know the \(\phi_{i,s}^{it}, \varphi_{j,s}^{it}\) variables values from the \(MM\) for every iteration, the right hand side of the routing equations becomes a datum.

Therefore, the \(SM^{it}\) can be reformulated as it was exposed in Cadarso and Marín [33], where the authors developed an integer model to determine train sequences once the RS assignment was known. In this case, for each iteration we will know the RS assignment from the \(MM\), so we can reformulate the \(SM\) into the mentioned formulation.

In the reformulated \(SM\) \((8.22)-(8.25)\) indexes \(i', j'\) are appearing. However, they do not have the same meaning as before. In the previous formulation \(i, j\) only referred to the type of operation (services, aggregations or disaggregations). However, \(i', j'\) are now numbering operations. For example, a service may be numbered as operation number 56: departure and arrival stations and times as well as assigned composition are associated to this operation number. Consequently, we can know whether two different operations are
compatible or not by only the operation number. This compatibleness is showed by the
new set $CO_{i'}$, which elements represent whether operations $i', j'$ are compatible or not.

Therefore, every time we are writing the index $i'$ in the reformulated $SM$, we are
actually referring to $i, s, c$. The same applies for $j'$. So the relationship between both
formulations may be summarized as:

$$\alpha_{i'} = \tilde{\alpha}_{it} \phi_{i,s}^c; \quad \beta_{j'} = \tilde{\beta}_{jc} \phi_{j,s}^c, \forall i, s, c \in I, S, C \text{ for each iteration } it.$$  

Similarly, dual variables $\kappa_{i'}$, $\lambda_{j'}$ are related to $\kappa_{i,s}^c, \lambda_{j,s}^c$.

$$\min \sum_{i' \in I} \sum_{j' \in I} \left( \psi_{pd} j' + \zeta_{cr} j' \right) seq_{i'}^{j'} \tag{8.22}$$

Subject to:

$$\sum_{j' \in CO_{i'}} seq_{i'}^{j'} = \alpha_{i'}^{it} \quad (\kappa_{i'}) \quad \forall i' \in I \tag{8.23}$$

$$\sum_{i' \in CO_{j'}} seq_{i'}^{j'} = \beta_{j'}^{it} \quad (\lambda_{j'}) \quad \forall j' \in I \tag{8.24}$$

$$seq_{i'}^{j'} \in \mathbb{R}^+ \quad \forall i', j' \in I \tag{8.25}$$

We have relaxed the integrality property of the binary variable $seq_{i'}^{j'}$ considering that
the relaxed $SM$ with integer data has an integer solution (see Theorem 8.5.1). The TR
model formulation used in this work was developed in order to provide feasible sequences
whatever the RS assignment is. This is achieved by ensuring minimum times between
operations. Hence, the routing $SM$ will always be feasible and no feasibility cuts are
needed for the $MM$.

The dual model of the reformulated $SM$ is as follows:

$$\max z = \sum_{i' \in I} \alpha_{i'} \kappa_{i'} + \sum_{j' \in I} \beta_{j'} \lambda_{j'} \tag{8.26}$$

$$\kappa_{i'} + \lambda_{j'} \leq \psi_{pd} j' + \zeta_{cr} j' \quad (seq_{i'}^{j'}) \quad \forall i', j' \in CO_{i'} \tag{8.27}$$

$$\kappa_{i'} \in \mathbb{R} \quad \forall i' \in I \tag{8.28}$$

$$\lambda_{j'} \in \mathbb{R} \quad \forall j' \in I \tag{8.29}$$

**Theorem 8.5.1** The linear programming relaxation $\min \{z(x) : A \cdot x = b, x \in \mathbb{R}^+\}$ of the
**TR problem** \( \min \{ z(\mathbf{x}) : A \cdot \mathbf{x} = b, \mathbf{x} \in \mathbb{Z}^+ \} \) **with integer data** \( b \) will have an optimal solution that is integer.

**Proof.** Suppose we have an optimal basis \( B \) from the reformulated \( SM \). From linear programming we know that \( B \) is a non singular submatrix of \( A \) where \( A \) is the coefficient matrix. Therefore, the optimal basis will be composed of \( \{0, 1\} \). Then, the first condition for \( B \) being totally unimodular (Wolsey [127]) is matched. Moreover, due to the problem characteristics, \( B \) will be always totally unimodular because feasibility mandates that every operation \( j \) must be preceded by a unique one, except for aggregations where at most two different operations must precede.

Hence, the optimal basis \( B \) will always be totally unimodular and \( \det(B) = \pm 1 \) (\( B \) is an optimal basis: \( \det(B) \neq 0 \)), so the linear relaxation solves the integer problem (Wolsey [127]).

In order to build the Benders optimality cuts, once we know the dual variables \( \kappa^c_i', \lambda^c_j' \) we must do a mapping to the variables in the \( MM \): \( \kappa^c_{i,s}', \lambda^c_{j,s}' \).

### 8.5.2 Benders Master Model

The Benders \( MM^u \) will be as follows:

\[
\begin{align*}
\min z &= \sum_{\ell \in L} \sum_{c \in C} oc_c \cdot km_{\ell,c} x^c_{\ell} + \sum_{\ell \in L^t} \sum_{a \in A_{\ell}} dpc_{a,\ell} dp_{a,\ell} + \\
&+ \sum_{s \in SC} \sum_{c,c' \in C} \vartheta_s \cdot \epsilon_{c,c'}^s + \omega \tag{8.30} \\
\sum_{c \in C} cap_c x^c_{\ell} &\geq pf_{a,\ell} - dp_{a,\ell} \quad \forall \ell \in L^t, a \in A_{\ell} \tag{8.31} \\
\sum_{c \in C} x^c_{\ell} &\cdot c = 1 \quad \forall \ell \in L^t \tag{8.32} \\
\sum_{c \in C} x^c_{\ell} &\cdot e \leq 1 \quad \forall \ell \in L^e \tag{8.33} \\
y^c_{s-} + \sum_{\ell \in L} \sum_{\alpha_{t,s}=1} x^c_{\ell} + \sum_{c' \in C} \epsilon_{s-c}^c + \sum_{c' \in C} cn_{c,c'} \cdot \delta_{c-c'}^c &= \\
y^c_{s+} + \sum_{\ell \in L} \sum_{\alpha_{t,s}=-1} x^c_{\ell} + \sum_{c' \in C} cn_{c,c'} \cdot \epsilon_{s-c'}^c + \sum_{c' \in C} \delta_{c-c'}^c \quad \forall s \in SC, c \in C \tag{8.34}
\end{align*}
\]
\[
\sum_{s \in SC} \sum_{c \in C} tu_c y^c_s + \sum_{\ell \in L} \sum_{c \in C_m} tu_c \beta_{\ell,c} x^c_{\ell} + \\
\sum_{s \in SC} \sum_{s' \in SC} \sum_{c, c' \in C} \mu_{s',s} \left( tu_c \epsilon^{c,c'}_{s} + tu_c \delta^{c,c'}_{s} \right) \leq \chi_m \quad \forall m \in M \tag{8.35}
\]

\[
\sum_{c \in C} tu_c y^c_s + \sum_{s' \in SC} \sum_{c \in C} \mu_{s',s} \left( tu_c \epsilon^{c,c'}_{s} + tu_c \delta^{c,c'}_{s} \right) \leq cap_s \quad \forall s \in SC \tag{8.36}
\]

\[
cc^{c,c'}_s = \epsilon^{c,c'}_s + \delta^{c,c'}_s \quad \forall s \in SC, c, c' \in C \tag{8.37}
\]

\[
y^c_{si} = y^c_{sf} \quad \forall s \in SC, c \in C \tag{8.38}
\]

\[
x^c_{\ell} = \varphi^c_{1,s} \quad \forall \ell \in L, c \in C, s \in SC : \alpha_{l,s} = 1 \tag{8.39}
\]

\[
x^c_{\ell} = \phi^c_{1,s} \quad \forall \ell \in L, c \in C, s \in SC : \alpha_{l,s} = 1 \tag{8.40}
\]

\[
\epsilon^{c,c'}_s = \varphi^{c,c'}_{2,s} \quad \forall s \in SC, c, c' \in M, C \tag{8.41}
\]

\[
\epsilon^{c,c'}_s = \phi^{c,c'}_{2,s+\epsilon} \quad \forall s \in SC, c, c' \in C \tag{8.42}
\]

\[
\delta^{c,c'}_s = \varphi^{c,c'}_{3,s} \quad \forall s \in SC, c, c' \in C \tag{8.43}
\]

\[
\delta^{c,c'}_s = \phi^{c,c'}_{3,s+\delta} \quad \forall s \in SC, c, c' \in C \tag{8.44}
\]

\[
\omega \geq \sum_{i \in I} \sum_{s \in S} \sum_{c \in C} \alpha_{i,c} \phi^{c,s}_{i,s} \kappa^{c,\ell}_{i,s} + \sum_{j \in I} \sum_{s \in S} \sum_{c \in C} \beta_{j,c} \varphi^{c,s}_{j,s} \lambda^{c,\ell}_{j,s} \quad \forall \ell \in AOBC_{it} \tag{8.45}
\]

where (8.45) are Active Optimality Benders Cuts at iteration \(it\) (AOBC\(_{it}\)).

### 8.6 Computational Experiments

All of our computational experience is for realistic cases drawn from RENFE's regional network in Madrid, also known as "Cercanías Madrid".

Our runs are performed on a Personal Computer with an Intel Core2 Quad Q9950 CPU at 2.83 GHz and 8 GB of RAM, running under Windows 7 64Bit, and our programs were implemented in GAMS/Cplex 12.

#### 8.6.1 Polishing Sequentially Obtained Solutions

Nowadays, RENFE makes its planning in a sequential manner. That is, the circulations are obtained by solving two problems in an independent way. The two problems are the rolling stock assignment and train routing problems. Once they are solved they obtain train circulations.
RENFE does not use operation research techniques to obtain the circulations. Consequently, their schedule may be improved by using them. See Cadarso and Marín [34] and Cadarso and Marín [33] for further details in the application of operations research techniques to RENFE planning. The authors proposed a sequential planning. However, sequential planning may lead to a suboptimal schedule. Consequently, we show in this section how integrated planning may produce better, smoother and more robust plans.

In the proposed Benders MM (8.30)-(8.45) we have optimality cuts (8.45). In these cuts, two different terms are appearing on the right hand side: the first one representing ending operations and the second one beginning operations. Applying Benders Decomposition with $AOBC_{it}$, we have observed that the algorithm needs a lot of iterations to provide acceptable solutions: during the initial iterations it provides solutions with an unaffordable number of composition changes and empty services. Therefore, the algorithm runs many iterations without providing any improvement with respect to sequentially obtained solutions, making the Benders MM bigger and more difficult to solve due to the large number of Benders optimality cuts.

The issue is to design a smart schedule to achieve robust sequences. However, the schedule for a train service is already fixed. Hence, it makes no sense to try to change it. Nevertheless, the schedule for empty services and shunting operations is not fixed, we may decide it. Therefore, we will only include in the optimality cuts those terms referring to empty trains and shunting operations.

Therefore, a new set $I'$ is defined. It is the operations set composed of empty trains, train units aggregations and train units disaggregations. So, the new Benders based Heuristic $AOBC_{it}$ ($HAOBC_{it}$) are showed in (8.46).

\[
\omega \geq \sum_{i \in I'} \sum_{s \in S} \sum_{c \in C} \tilde{c}_{i,c} \phi_{i,s} - \tilde{c}_{i,t} + \sum_{j \in I'} \sum_{s \in S} \sum_{c \in C} \tilde{d}_{j,c} \varphi_{j,s} - \lambda_{j,s} \forall it \in HAOBC_{it} \tag{8.46}
\]

The Benders heuristic consists of obtaining an initial RS assignment. Then, the train routing is obtained, and based on that routing a new RS assignment is obtained. Consequently, RS assignment and shunting schedules change every iteration. For the first iteration we provide to the Benders heuristic with a feasible rolling stock assignment problem solution.

The proposed heuristic is not an exact method: we are not obtaining exact solutions to the integrated model but improving sequentially obtained results. Therefore, a stopping criterion must be selected. For these study cases this criterion will be as follows: the
Table 8.1 Line C5 Benders based heuristic solution at each iteration

<table>
<thead>
<tr>
<th>Iteration</th>
<th>#C</th>
<th>TSOC</th>
<th>ESOC</th>
<th>DPC</th>
<th>#CC</th>
<th>EDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64</td>
<td>80099.7</td>
<td>1265</td>
<td>3554</td>
<td>20</td>
<td>111.8</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>80563.44</td>
<td>1645.3</td>
<td>3248</td>
<td>14</td>
<td>72.05</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>80570.64</td>
<td>1393.04</td>
<td>3248</td>
<td>16</td>
<td>75.01</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>80570.64</td>
<td>1490.24</td>
<td>3248</td>
<td>14</td>
<td>67.16</td>
</tr>
<tr>
<td>5</td>
<td>66</td>
<td>80727.6</td>
<td>1359.04</td>
<td>3248</td>
<td>14</td>
<td>68.69</td>
</tr>
<tr>
<td>6</td>
<td>66</td>
<td>80346</td>
<td>1553.92</td>
<td>3639</td>
<td>10</td>
<td>69.76</td>
</tr>
<tr>
<td>7</td>
<td>66</td>
<td>80413.68</td>
<td>1457.44</td>
<td>3554</td>
<td>14</td>
<td>61.03</td>
</tr>
<tr>
<td>8</td>
<td>66</td>
<td>80346</td>
<td>1526.32</td>
<td>3639</td>
<td>12</td>
<td>71.89</td>
</tr>
<tr>
<td>9</td>
<td>66</td>
<td>80346</td>
<td>1563.12</td>
<td>3639</td>
<td>10</td>
<td>69.54</td>
</tr>
<tr>
<td>10</td>
<td>66</td>
<td>80360.4</td>
<td>1475.2</td>
<td>3696</td>
<td>12</td>
<td>66.07</td>
</tr>
<tr>
<td>11</td>
<td>66</td>
<td>80346</td>
<td>1574.32</td>
<td>3639</td>
<td>10</td>
<td>63.53</td>
</tr>
<tr>
<td>12</td>
<td>66</td>
<td>80742</td>
<td>1612.1</td>
<td>3326</td>
<td>12</td>
<td>70.48</td>
</tr>
</tbody>
</table>

The heuristic will terminate if five consecutive iterations do not improve the best found solution up to that iteration. We define the best solution as that one with the lowest expected propagated delay.

8.6.2 Line C5

We can see the evolution of the solution for each of the performed iterations in Table 8.1 for line C5 (Figure 5.1). In the first column the iteration number is shown. The number of train units used in the proposed solution (#C) in the second column, the train service operating costs (TSOC) in the third column, the empty service operating costs (ESOC) in the fourth column, disrupted passenger costs (DPC) are shown in the fifth column, the number of composition changes (#CC) in the sixth column and the expected delay propagation (EDP) in minutes in the last column.

In each iteration a different solution is obtained. It may seem that some solutions are equal in some costs. However, we must account for the fact that the empty trains and shunting schedule is being changed. Consequently, operating costs might be equal but the sequences are being changed and so does the EDP. Moreover, each iteration provides the operator with a feasible planning which may be implemented. The operator may decide to choose among them according to several criteria such as train units maintenance and crew scheduling. Therefore, this approach provides different solutions to the operator.

The obtained solution is summarized in Table 8.2. In the second column #C is showed, which is slightly reduced. TSOC are in the third column, ESOC in the fourth one, DPC in the fifth column, #CC in the sixth one, EDP in the seventh column, and the expected
delay reduction (EDR) in the last one. The Benders based heuristic chooses the solution that produces the lowest EDP. We can see how the obtained solution in the integrated approach (Benders Heuristic) solved with the proposed heuristic improves the EDP with respect both the current solution operated by RENFE (Current solution) and the solution obtained by the sequential approach (RS & TR solution, see Cadarso and Marín [34] and Cadarso and Marín [33]), and reduces the number of performed composition changes with respect the solution obtained by the sequential approach. This reduction in the number of composition changes is always good news because of the possibility of malfunction: the lower the number of composition changes, the greater of the robustness degree. However, the achievement of this robustness is not for free. We must mention that TSOC and ESOC increase a bit compared to the sequential approach. However, a great EDR compared to the current an sequential solutions is obtained. In order to get a deeper insight in the expected delay reduction, the distribution of the propagated delay is shown in Figure 8.1 for the integrated and sequential approaches. In the integrated approach the number of operations connections (defined by two different operations performed by the same train unit) (vertical axis) with positive expected propagated delay (horizontal axis, expressed in minutes) is slightly reduced. Consequently, the number of needed swapping operations will be also reduced. Therefore, the integrated planning turns out to be more robust and smoother to be operated.

![Figure 8.1 Line C5: number of operations connections for each propagated delay in the integrated approach (left side) and in the sequential approach (right side)](image)

In the integrated approach the computational time is 1620 seconds. It is higher than in the sequential approach. However, the obtained solution has fewer composition changes than the solution in the sequential approach. Therefore, the solution is more robust. Moreover, the are fewer operations connections with positive expected propagated delay. These issues joint with the fact that the computational is still reasonable for the planning
Table 8.2 Line C5: Benders heuristic solution, sequential (RS & TR) solution and current solution

<table>
<thead>
<tr>
<th></th>
<th>#C</th>
<th>TSOC</th>
<th>ESOC</th>
<th>DPC</th>
<th>#CC</th>
<th>EDP</th>
<th>EDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benders Heuristic</td>
<td>66</td>
<td>80413.68</td>
<td>1457.44</td>
<td>3554</td>
<td>14</td>
<td>61.03</td>
<td>56.9%</td>
</tr>
<tr>
<td>RS &amp; TR solution</td>
<td>64</td>
<td>80099.7</td>
<td>1265</td>
<td>3554</td>
<td>20</td>
<td>111.8</td>
<td>20.91%</td>
</tr>
<tr>
<td>Current solution</td>
<td>74</td>
<td>109765.2</td>
<td>2232.1</td>
<td>874</td>
<td>0</td>
<td>141.37</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 8.3 Lines C3&C4 Benders based heuristic solution at each iteration

<table>
<thead>
<tr>
<th>Iteration</th>
<th>#C</th>
<th>TSOC</th>
<th>ESOC</th>
<th>DPC</th>
<th>#CC</th>
<th>EDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63</td>
<td>87837.52</td>
<td>4090.83</td>
<td>2177</td>
<td>34</td>
<td>163.98</td>
</tr>
<tr>
<td>2</td>
<td>62</td>
<td>88937.23</td>
<td>3810.84</td>
<td>2177</td>
<td>38</td>
<td>179.2</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>89811.43</td>
<td>4624.2</td>
<td>3403</td>
<td>32</td>
<td>152.25</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>88653.73</td>
<td>3796.80</td>
<td>2723</td>
<td>32</td>
<td>147.95</td>
</tr>
<tr>
<td>5</td>
<td>66</td>
<td>90206.71</td>
<td>4179</td>
<td>2177</td>
<td>36</td>
<td>164.46</td>
</tr>
<tr>
<td>6</td>
<td>65</td>
<td>90532.66</td>
<td>3780.6</td>
<td>2077</td>
<td>40</td>
<td>183</td>
</tr>
<tr>
<td>7</td>
<td>65</td>
<td>90128.2</td>
<td>4061.61</td>
<td>2177</td>
<td>38</td>
<td>182.37</td>
</tr>
<tr>
<td>8</td>
<td>66</td>
<td>90754.33</td>
<td>3608.19</td>
<td>2177</td>
<td>38</td>
<td>182.37</td>
</tr>
</tbody>
</table>

horizon we are working on makes the integrated approach interesting for the operator.

8.6.3 Lines C3&C4

Again, we use the proposed Benders heuristic to solve the integrated model for lines C3 and C4 (Figures 5.2 and 5.3). The results obtained in this study case lead us to similar conclusions as that ones presented for the line C5 study case.

In Table 8.3 we can see the evolution of the solution for each of the performed iterations. In the first column the iteration number is shown. The #C in the second column, the TSOC in the third column, the ESOC in the fourth column, DPC are shown in the fifth column, the #CC in the sixth column and the EDP in minutes in the last column. In each iteration a different solution is obtained. Among all the feasible solutions provided by the heuristic we choose the solution with the lowest EDP.

The obtained solution is summarized in Table 8.4. Table 8.4 may be read as Table 8.2. We can see how the obtained solution by the integrated approach solved with the proposed heuristic improves the EDP with respect both the current solution operated by RENFE (Current solution) and the solution obtained by the sequential approach (RS & TR solution), and reduces the number of performed composition changes with respect the solution obtained by the sequential approach.

Again, in order to get a deeper insight in the EDR, the distribution of the propagated
Table 8.4 Lines C3&C4 heuristic solution vs. current solution

<table>
<thead>
<tr>
<th></th>
<th>#C</th>
<th>TSOC</th>
<th>ESOC</th>
<th>PEC</th>
<th>#CC</th>
<th>EDP</th>
<th>EDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benders Heuristic</td>
<td>64</td>
<td>88653.73</td>
<td>3796.80</td>
<td>2723</td>
<td>32</td>
<td>147.96</td>
<td>24.4 %</td>
</tr>
<tr>
<td>RS &amp; TR solution</td>
<td>63</td>
<td>87837.52</td>
<td>4090.83</td>
<td>2177</td>
<td>34</td>
<td>163.98</td>
<td>16.19 %</td>
</tr>
<tr>
<td>Current solution</td>
<td>72</td>
<td>136633.86</td>
<td>6083.88</td>
<td>1550</td>
<td>0</td>
<td>195.65</td>
<td>-</td>
</tr>
</tbody>
</table>

delay is shown in Figure 8.2 for the integrated and sequential approaches. In the integrated approach the number of operations connections (vertical axis) with positive expected propagated delay (horizontal axis, expressed in minutes) is reduced. Consequently, the number of needed swapping operations will be also reduced.

The computational time is increased compared to the line C5. Now, the time needed to solve the model is 3605 seconds. This is due to the fact that in this case two different lines are being solved, there are more depot stations and possibilities for shunting operations are greater. Thus, Benders heuristic’s cuts are larger and become the problem more difficult to solve. Nevertheless, as the purpose of this work is to propose the schedule well in advance to the execution of it, the computational time required is still reasonable.

![Figure 8.2 Line C3&C4: number of operations connections for each propagated delay in the integrated approach (left side) and in the sequential approach (right side)](image)

8.7 Summary

We have proposed a new approach to solve the rolling stock circulations. In order to obtain them two different problems must be addressed: the rolling stock assignment problem and the train routing problem.

The integrated approach is a good frame to improve the robustness degree of the system. However, the proposed integrated model to solve the rolling stock circulations
has an enormous size to be solved by the current commercial software. Therefore, we propose a Benders based heuristic to solve the proposed model.

Computational experiments show how the current solution operated by RENFE can be improved: more robust and smoother solutions are obtained. RENFE planners do not use operations research techniques for planning purposes. The proposed integrated approach improves the solution provided by RENFE planners and the solutions obtained using operations research techniques in a sequential fashion. We are able to produce plans with fewer composition changes which are considered to be dangerous by planners. We have also reduced the number of operation connections with positive expected propagated delay. Consequently, the number of re-scheduling operations (i.e., swapping operations) are reduced making easier the recovery to the operator.
Chapter 9

RECOVERY OF DISRUPTIONS IN RAPID TRANSIT NETWORKS

This chapter studies the disruption management problem of rapid transit rail networks. Besides optimizing the timetable and the rolling stock schedules, we explicitly deal with the effects of the disruption on the passenger demand.

We propose a two-step approach that combines an integrated optimization model (for the timetable and rolling stock) with a multinomial logit model (for the passengers’ behavior).

We report our computational tests on realistic problem instances of the Spanish rail operator RENFE. The proposed approach is able find solutions with a very good balance between various managerial goals within a few minutes.

9.1 Introduction

During the daily operations of a dense railway network, incidents may cause the railway traffic to deviate from the planned operations. These incidents may make it impossible to operate the schedule as it was originally planned. In such a situation the operator needs to adjust the timetable and the rolling stock assignment for the time interval of the incident, and to carry out further recovery steps in order to get back to the original schedules.

The first task after noticing an incident is to determine whether or not it requires a substantial active intervention. If it does, the railway operations are said to be disrupted, and plans must be designed in order to recover from the disrupted situation. Disruptions may be caused by infrastructure blockage, failing rolling stock, and crew shortage.
Regardless of the cause of a disruption, it has an impact on the railway system. The impact is generally in the form of a change in the system settings, a change in resource availability, or both. Disruptions usually involve a change in resource availability. The response to a change in resource availability is to replan the current operations to apply only the available resources which may include giving up some of the planned services. A disruption may also cause a change in the system settings. Closing a station (or part thereof) temporarily is an example of a change in the system settings that affects the system’s ability to operate. A further change in the system environment is a deviation in demand because the passengers are free to choose their own path in the network.

A particularly challenging aspect of disruption management is the fact that the passenger demand is influenced by the operator’s recovery actions, too. Timetabling decisions may cancel traveling options, or insert additional ones, while the rolling stock decisions influence the capacity of the trains. The actual passenger flows emerge from the interaction of the timetable, the rolling stock schedule and the passengers.

In the current railway practice, the recovery problem is solved in a sequential manner. When a disruption occurred, first a new timetable is computed accounting for rolling stock availability; as a result, the rolling stock units will not finish their daily duties at the location where they were planned to: to avoid deadheading trips, the rolling stock schedules are modified such that the rolling stock is balanced before the end of the day. However, this approach has limitations: computing a new timetable without accounting for rolling stock may produce a suboptimal timetable or even an infeasible one for rolling stock assignment purposes. Therefore, we are interested in an integrated approach to obtain an optimal solution for timetable and rolling stock assignment accounting for demand rerouting.

The disruption management process has several objectives. The first goal is to provide the best possible service quality by accommodating the largest possible part of passenger demand. The second goal aims at easing the rescheduling process itself by minimizing the differences between the original (undisrupted) plan and the recovery plan. Third, the operators often want to quickly return to the original plan once the disruption is over. That is, the duration of the recovery period is to be minimized.

This chapter is organized as follows. In Section 9.2 a literature overview is given. Section 9.3 describes the problem in detail. Section 9.4 is devoted to the mathematical model. In Section 9.5 we present our computational experiments.
9.2 State of Art

Jespersen-Groth et al. [78] deal with disruption management in passenger railway transportation. They describe the disruption management process and the roles of the different actors involved in it. Furthermore, they discuss the three main subproblems in railway disruption management: timetable adjustment, and rolling stock and crew re-scheduling. De Almeida et al. [56] propose an approach for dealing with large scale disruptions where track capacity is greatly reduced. As once a disruption has occurred, the first dispatching task is to keep the railway system running, the first decisions are taken under extreme time pressure. Therefore, decision proposals should be generated quickly. Thus, they propose a heuristic approach to re-building a passenger transportation plan in real time. Kroon and Huisman [84] also state that, in case of a disruption, rescheduling is a time critical situation where every minute counts. They describe models and algorithms for real-time rolling stock rescheduling and real-time crew rescheduling. Mesa et al. [96] study rescheduling of train services, reducing the current supply along one transportation line in order to reinforce the service of another line, exploited by the same public operator, which has suffered an incident. A methodology, based on a geometric representation of solutions which allows the use of discrete optimization techniques, is developed.

During a disruption, the dispatchers try to use all available rolling stock to transport as many passengers as possible in the right direction. As a result, the rolling stock units will not finish their daily duties at the location where they were planned to. Budai et al. [28] state that in order to prevent expensive deadheading trips, it is attractive to modify the rolling stock schedules such that the rolling stock is balanced before the night.

Nielsen [101] studies the rescheduling of passenger railway rolling stock in disruption management and in the short-term planning phase. He formalizes the rolling stock rescheduling problem as a heuristic approach that uses three steps. The first step creates a rolling stock schedule based on the current timetable and passenger demand. The second step simulates the passenger response, and the third step interprets the passenger response. All test instances are based on the major Dutch railway operator NS.

Walker et al. [124] are among the first ones to deal with the integration of timetabling and resource scheduling in disruption management. They present a model that manipulates the timetable and the crew schedule at the same time. The objective is to simultaneously minimize the deviation of the new timetable from the original one, and the cost of the crew schedule.

In the airline industry, the recoverability of the system from disruptions has been studied deeper. Clausen [49] gives a short overview over the methods used for planning
and disruption management in the airline industry. Then he describes and discusses the situation regarding railway optimization.

It is also current practice to determine recovery plans in a primarily sequential manner, first recovering aircraft, then crew, and then passengers (Filar et al. 62). With respect to aircraft recovery, Jarrah et al. 77 consider the aircraft schedule recovery problem and propose two network models to address aircraft shortages. The objective of the first model is to determine flight leg departure times that minimize total flight delay costs, and the second is to select flight leg cancellations that minimize cancellation costs. Thengvall et al. 119 extend the approach of Jarrah et al. 77 to consider flight leg departure scheduling and cancellations simultaneously. Rosenberger et al. 106 present an optimization model that reschedules legs and reroutes aircraft by minimizing an objective function involving rerouting and cancellation costs. Then, they revise the model to minimize crew and passenger disruptions. Stojković et al. 115 propose a model to select flight leg departure times, considering crew transfers, rest periods, passenger connections, and aircraft maintenance, but not including cancellation decisions.

Bratu and Barnhart 26 present airline schedule recovery models and algorithms that simultaneously develop recovery plans for aircraft, crews, and passengers by determining which flight leg departures to postpone and which to cancel. The objective is to minimize jointly airline operating costs and estimated passenger delay and disruption costs. Dumas and Soumis 59 and Dumas et al. 60 propose a framework for revenue management where aircraft scheduling is combined with a passenger flow model.

9.3 Problem Description

In this section, the recoverability problem in rapid transit networks is described in detail.

9.3.1 The Disruption

In this chapter, we focus on the common type of disruptions when a line segment between two neighboring stations becomes fully or partially blocked for a certain time period. The partial blockage is particularly interesting in that two-way traffic needs to be scheduled on a single pair of rails.

The impact of this disruption will be a change in the network topology and in the resource availability. A completely new train schedule is needed because the planned operations are infeasible in the new scenario. This issue will cause a demand deviation because some part of the demand will not be able to realize its travel as it was planned.
As we are studying a real life problem, the railway infrastructure is not isolated from other modes of transport. We will consider the existence of the Metro network. This Metro network has several stations in common with the rapid transit railway network. However, they are independent, they use different infrastructure and they are operated by different operators. Therefore, when a disruption occurs passengers may find an attractive path using both, the railway and Metro network.

As we have explained above, the disruption will change the network settings in such a way that infrastructure capacity and resource availability is reduced. To be more concise the infrastructure capacity that will be reduced is the one referring to some arc or arcs in the network. This capacity availability will be denoted by the parameter $a_{in_t}$, its value indicating the number of pairs of rails that can be used in time period $t$. The value $a_{in_t}$ is thus 0 if the arc is fully blocked; the value is 1 if the arc is partially block, and the value is 2 if the arc has its full capacity.

### 9.3.2 Timetable

We distinguish two types of train services: the planned train services represented by $\ell \in L^p$ and the emergency services represented by $\ell \in L^e$. The former are the trains scheduled for a normal (undisrupted) situation; emergency trains are inserted to the schedule during the disruption in order to alleviate its negative effects in passengers. We will refer to the set of all services as $L = L^p \cup L^e$.

Planned services may be canceled due to some disruption. We will not consider the possibility of changing the planned train services’ departure times on a time window. After all, the frequencies in a rapid transit network are rather close to their maximum value during rush hours; this maximum value is limited by the headway time which is imposed by the infrastructure manager.

For emergency services $\ell \in L^e$ the model will decide whether they are used or not. An emergency service represents a feasible movement between depot stations, and it is characterized by a departure station, an arrival station, every intermediate arc and the departure time. We define a feasible movement as a physical movement in the network once the disruption has started. From this point, the model decides whether an emergency service is assigned to a departure time or not.

For planned and emergency train services the headway must be maintained in every infrastructure they come through.
9.3.3 Passengers

Once the disruption has occurred, passengers will have to use the new network topology to reach their destination. First, they will have to find a path in the modified network, then wait for a train service and finally enter the train if enough capacity is available. Passengers who cannot enter a train due to lacking capacity are willing to wait for, say, 10 minutes, and try to board the next train; otherwise, the passenger is supposed to leave the system and use another means of transport. Train services and its rolling stock will be decided in an integrated way accounting for the expected passenger groups decisions.

**Passengers Groups.** The demand is characterized by an origin, a destination and a departure time. This information may be represented by passenger group \( w = (o, d, \tau) \), where \( o \in S \) is the departure station, \( d \in S \) the arrival station, \( \tau \in T \) the desired departure time and \( w \in W \) the set of passengers groups. The size of the passenger group is denoted by \( g_w \).

**Paths.** The demand will be realized through available paths \( p \in P \) in the network. Each passenger group \( w \in W \) will be able to choose a path \( p \in P_w \), where \( P_w \subset P \) denotes the set of paths attending \( w \in W \). Passengers within the same passenger group may travel by different paths, that is, passenger groups may be split.

As we are working in a rapid transit system, where different modes of transportation exist, we also will include paths containing these alternative modes. For example, we could have a path composed of some arcs in the railway network, and some arcs in the Metro network because they are interconnected in some stations. Moreover, we could also have paths composed of different lines in the railway network.

Each path is characterized by its origin, destination, the arcs belonging to it and its expected travel time. The total expected travel time will be the sum of the on-board time, transfer time and waiting time. The demand will choose its path based on the expected travel time.

**Passengers’ reaction to the disruption.** We propose a way to anticipate passenger demand before computing the resource schedules; the anticipated demand is used to guide the integrated optimization model for the timetabling and rolling stock scheduling.

The anticipated demand is based on a number of assumptions. First, we assume that the same passenger groups show up as on a normal, undisrupted day. Second, we assume that the passengers choose their travel path according to the multinomial logit model (see Section 9.3.4). Note that the travel path is purely geographic at this point since the timetable is not known yet. Third, we assume that the passengers stick to their path choice, even if the realized travel time becomes much higher than what they expected.
The travel times are estimated as follows. The exact timetable may not yet be known, but the passengers can rely on the train frequencies, e.g., a train every 10 minutes. Then they can expect an average waiting time of 5 minutes. As for the partially blocked arc, they can reasonably assume that trains will run alternatingly left-to-right and right-to-left, yielding estimated travel time from one end of the disrupted arc to another. Having chosen a path, the passenger appears in the passenger demand on each arc along the travel path.

We also note that per-trip passenger demand is impractical in railway network with high train frequencies. The passengers can be expected to wait for the next train if they cannot embark an overcrowded train. Therefore we actually calculate the demand of an arc for time intervals of, say, 10 minutes. The rolling stock assignment model of Section 9.4 compares this demand to the sum of the train capacities during the given time interval.

The proposed model for the passenger demand is valid as long as each passenger is accommodated in the trains. However, if a passenger cannot take a train (due to insufficient capacity), his/her presence as demand on later trips becomes meaningless. Our optimization model cannot cope with this issue, therefore the outcome of our model needs an afterwards validation and discussion.

9.3.4 Multinomial Logit Model

As we have said above, a disruption will change the system settings. Due to the new settings and according to the explained assumptions regarding passenger groups, passengers will choose a new path under these new settings. These choices will depend on the new schedule we are obtaining in an integrated way accounting for the timetable and rolling stock resources at the same time.

Discrete choice models have played an important role in transportation modeling for the last years. These models consider that the demand is the result of several decisions of each individual in the population under consideration. These decisions usually consist of a choice made among a finite set of alternatives (Ben-Akiva and Lerman [17], Ortúzar [103]).

The multinomial logit model allows us to capture how individuals are making choices. We must define the decision-maker and his/her characteristics, the alternatives as the possible options of the decision-maker, the attributes of each potential alternative the decision-maker is accounting for, and the decision rules describing the rules used by the decision-maker.

The utility of chosen path is a function of the attributes of the alternative itself and
of the decision-maker. The deterministic part of the utility that a decision-maker is associating with alternative \( p \in P \) is:

\[
v^w_p = v(a^w_p) \quad \text{for} \quad w \in W, p \in P_w
\]

(9.1)

where \( a^w_p \) is a vector containing all attributes of alternative \( p \in P \) for each passenger group \( w \in W \).

The utility function for every path is calculated as the sum of different terms: the traveling time (as the sum of the travel time of each of link of the path), the transfer time and the waiting time. That is,

\[
v^w_p = \beta_1 \ot^w_p + \beta_2 \tt^w_p + \beta_3 \wt^w_p \quad \text{for} \quad p \in P_w, \forall w \in W
\]

(9.2)

where \( \ot^w_p \) is the on-board time, \( \tt^w_p \) is the transfer time and \( \wt^w_p \) is the waiting time for each path \( p \) attending demand \( w \). \( \beta_1, \beta_2, \beta_3 \) represent the utility value of each of these different kinds of times.

Supposing that the utility function’s error terms are (i) independently distributed, (ii) identically distributed, and (iii) Gumbel distributed with a location parameter and a scale parameter \( \theta > 0 \), then the probability of choosing a given itinerary \( p \) among the set \( P_w \) by the demand \( w \) will be as follows (Ben-Akiva and Lerman [17]):

\[
P(p \mid w) = \frac{e^{-\theta(\alpha^w_p + v^w_p)}}{\sum_{\pi \in P_w} e^{-\theta(\alpha^w_{\pi} + v^w_{\pi})}}
\]

(9.3)

where \( P(p \mid w) \) is the probability that passenger in group \( w \) chooses path \( p \). Here \( \alpha^w_\pi \) is a parameter that represents the relative attractiveness of path \( \pi \) to demand \( w \) due to factors not included in the utility function such as security, comfort, ticketing, on so on. The multinomial logit model parameters are provided by the operator. They are obtained based on the operator’s passengers counts and historical data fittings. The multinomial logit model has been calibrated by the operator and the obtained data are confidential.

Due to the huge number of possibilities for passengers groups to be realized, we consider the passengers’ flows in arcs only, instead of passengers’ flows in paths. In this way, we obtain reasonable computational times; recall that we are studying a problem where decisions must be made in a time horizon of minutes.

The per-path demand is transformed into per-arc demand as follows. Let \( a \) be an arc in the network, and let \( \tau \) be a time period in which the demand is to be measured. Then
the per-arc demand is computed by
\[ p_{f_{a, \tau}} = \sum_{w \in W} \sum_{p \in P_w} \delta_{w, p}^{a, \tau} \cdot P(p|w) \cdot g_w \] (9.4)

where \( \delta_{w, p}^{a, \tau} \in \{0, 1\} \) expresses whether or not passenger group \( w \) using path \( p \) is coming through arc \( a \) during time period \( \tau \). That is, we assume that each group splits according to the probabilities \( P(p|w) \), and we sum up these splits passenger groups on each arc. The values \( p_{f_{a, \tau}} \) express the demand in the integrated timetable and rolling stock optimization model.

9.4 Integrated Timetable and Rolling Stock Rescheduling Model

The Integrated Timetable and Rolling Stock Rescheduling Model (ITRSRM) is a Mixed Integer Linear Programming model. It aims at computing the timetable and the rolling stock schedule for a disrupted rapid transit network, and balances several objective criteria. The planning period contains both the time interval of the disruption itself and the recovery period which is needed to fully return to the original schedule.

The ITRSRM is based on the rolling stock model proposed by Cadarso and Marín [34]. They considered the railway rolling stock problem for rapid transit networks. ITRSRM also has strong similarities with the model proposed by Cadarso and Marín [36] where the authors develop a model to design the timetable and the rolling stock assignment. Compared to these two papers, the novelty of the current paper lies in the following aspects.

- The possibility of canceling planned train services is included.
- ITRSRM admits newly inserted, so called emergency train services. These are scheduled in order to alleviate the effects of the disruption on both passengers and rolling stock.
- The network topology may change along the planning period: the disrupted arc may admit one-way traffic, two-way traffic, or no traffic at all.

The model minimizes a combination of system-related and service-related criteria subject to constraints for the underlying timetabling and rolling stock scheduling problems. The purpose of the constraints is summarized as follows.
As for the timetable, headway times are enforced; emergency trains are inserted; and direction of the traffic on the disrupted arc is decided;

- The passenger demand is linked to the capacity of the allocated train units;
- As for the rolling stock, the amount of used rolling stock is limited; each trip gets a composition assigned; the storage and shunting capacity of the stations is controlled.

Our model treats the demand heuristically. The model is unable to trace individual passengers; instead, it considers demand on the arcs (i.e., between successive stations). Passengers on a longer journey appear in the demand of each arc underway. Whenever the demand of an arc exceeds the allocated capacity, part of the demand remains unsatisfied: these passengers are denied, and they are supposed to leave the system. However, the demand on successive arcs are not linked to each other. Therefore a denied passenger still shows up in the demand of later arcs. We discuss the justification for this heuristic demand treatment in Section 9.5.2.

In our model, the relationships between the data and variables are considered within a directed space-time graph, \( G(S, A) \), where \( S \) is the set of stations and \( A \) is the set of arcs. Each arc \( a \) is defined by \((s, t, s', t')\), where \( s \) and \( s' \) are the origin and destination nodes, \( t \) is the departure time, and \( t' \) is the arrival time. That is, \( t' = t + t_a \), where \( t_a \) is the time to move from \( s \) to \( s' \). It is assumed that this time is known and fixed for each arc. This means that in the ITRSRM, in which an arc is denoted by \( a \), this may be understood as \( a = (s, s', t) \).

### 9.4.1 Sets

In order to be able to formulate ITRSRM, we need to define the following sets.

- \( S \) is the set of stations.
- \( SC \subseteq S \) denotes the set of depot stations.
- \( A \) is the set of arcs. The arcs represent the infrastructure. Each arc \( a \in A \) is from departure station \( ds_a \in S \) to arrival station \( as_a \in S \).
- \( L \) set of train services.
- \( L^p \subset L \) set of planned train services.
- \( L^e \subset L \) set of emergency train services.
• $L_a \subseteq L$ is the set of train services that use arc $a \in A$.

• $T$ is the set of time intervals. A time interval $t \in T$ represents a certain interval in time, for example from 8:00 to 8:01.

• $TD_a \subseteq T$ is the set of time intervals through which the demand is counted in each arc $a$.

• $M$ is the train unit type set.

• $C$ is the compositions set.

• $C_m$ is the set of compositions made of train units type $m$.

• $IT \subset T$ is the set of time periods during which the incident is active.

• $LCS_{\ell,s}$ is the set of time periods during which train service $\ell \in L$ comes through station $s \in S$.

• $INO, ISO \subset A$ are the set of arcs belonging to a line segment between two neighboring stations which are affected by the disruption. The first set contains the arcs with a riding direction which is the opposite one to the riding direction in the second set in an undisturbed situation.

9.4.2 Variables

The most central decision variables are $x_{\ell,c} \in \{0,1\}$, defined for $\ell \in L, c \in C$. Their values indicate whether composition $c \in C$ is scheduled for service $\ell \in L$. Note that $x_{\ell,c}$ are the only variables that link the timetabling and passenger-related constraints to the rolling stock constraints. Therefore the proposed integrated model can be used for any underlying rolling stock scheduling problem as long as it is expressed in terms of $x_{\ell,c}$.

The model contains the following additional variables.

• $y_{\ell} \in \{0,1\}$, defined for $\ell \in L_p$, to indicate whether service $\ell \in L_p$ is canceled;

• $yt_{s,t}^c \in \mathbb{Z}^+$, defined for $s \in SC, t \in T, c \in C$, to denote the number of compositions $c$ in station $s$ at $t$ period.

• $dp_{a,\tau} \in \mathbb{Z}^+$, defined for $a \in A, \tau \in TD_a$, to denote the number of denied passengers due to insufficient capacity in each arc $a, \tau$;
• \( \rho_{c,s,t} \in \mathbb{Z}^+ \), defined for \( s \in SC, t \in Tc \in C \), to denote the number of rotations ending during \( t \) in depot \( s \) with composition \( c \); note that a rotation is always needed before a train can depart;

• \( em_{c,s,s',t} \in \{0,1\} \), defined for \( s, s' \in SC, t \in T, c \in C \), to indicate whether an empty movement is performed at \( t \) period from station \( s \) to \( s' \) with composition \( c \);

• \( \epsilon_{s,t} \in \{0,1\} \), defined for \( s \in SC, t \in T, c, c' \in C \), to indicate whether an aggregation is started during \( t \) in depot \( s \) from composition \( c \) to composition \( c' \);

• \( \delta \epsilon_{s,t} \in \{0,1\} \), defined for \( s \in SC, t \in T, c, c' \in C \), to indicate whether a disaggregation is started during \( t \) in depot \( s \) from composition \( c \) to composition \( c' \);

• \( \alpha_{t} \in \{0,1\} \), \( \beta_{t} \in \{0,1\} \), defined for \( t \in IT \), to indicate which riding direction is opened in the line segment between two neighboring stations affected by the disruption. If \( \alpha_{t} \) takes value 1, one of the riding direction is opened during \( t \) and the opposite direction is closed. Similarly for \( \beta_{t} \).

For the sake of clarity we declared all variables to be integral. We note that the nature of the constrains allows us to relax the integrality of \( yt_{c,s,t} \), \( \alpha_{c,s,t} \) and \( \rho_{c,s,t} \).

### 9.4.3 Objective Function

The objective function of the model reads as follows.

\[
\min z = \sum_{\ell \in L} \sum_{c \in C} oc_c km_\ell x_{\ell,c} + \\
\sum_{s,s' \in SC} \sum_{t \in T} \sum_{c \in C} oc_c km_{s,s',t} em_{s,s',t}^c + \sum_{s \in SC} \sum_{t \in T} \sum_{c,c' \in C} \vartheta_{s,t} \cdot cc_{c,c'} + \\
\sum_{\ell \in L^p} canc_\ell y_\ell + \sum_{a \in A} \sum_{\tau \in T} dpc_{a,\tau} dp_{a,\tau} + \\
\sum_{\ell \in L^p} \sum_{c \in C} \kappa_\ell \left| x_{\ell,c} - \hat{x}_{\ell,c} \right| + \sum_{s,s',t \in ST} \sum_{c \in C} \lambda_t \left| em_{s,s',t}^c - \hat{em}_{s,s',t}^c \right|
\]

The objective terms, in the given order, penalize the following quantities.

• Operating costs of planned and emergency services; here \( oc_c \) is the operating cost per kilometer and \( km_\ell \) is the distance in kilometers of service \( \ell \);
• Operating costs of empty movements; here \( km_{s,s'} \) is the distance in kilometers from \( s \) to \( s' \).

• Composition changes; here \( \vartheta_{s,t} \) is the cost of a composition change at depot \( s \) in time period \( t \).

• Cancellation of services; here \( canc_\ell \) is the cancellation cost for service \( \ell \).

• Denied passengers; here \( dpc_{a,\tau} \) is the cost per denied passenger due to insufficient capacity in each arc \( a \) during time period \( \tau \).

• Deviation from the schedule of commercial services; here \( \kappa_\ell \) is the penalty for changing the rolling stock assignment of a commercial service \( \ell \), while \( \hat{x}_{\ell,c} \) indicates the rolling stock assignment on a normal day.

• Deviation from the schedule of the empty movements; here \( \lambda_t \) is the penalty for changing the rolling stock assignment of an empty movement \( t \), while \( \hat{em}_{s,s',t} \) indicates the rolling stock assignment on a normal day.

The intuitive formulation of the last two terms is non-linear; however, the binary character of the variables \( x_{\ell,c} \) and \( em_{s,s',t} \) admits a straightforward linearization. We also note that the penalty values \( \kappa_\ell \) and \( \lambda_t \) are increasing with time. Therefore the last two terms of the objective attempt to minimize the length of the recovery period.

The objective function does not minimize passengers travel time because they choose the path according to their behavior which is modeled in the multinomial logit model.

### 9.4.4 Timetabling Constraints

The first set of timetabling constraints enforces the headway requirements.

\[
\sum_{\ell \in L} \sum_{t_1 \in LCS_{\ell,s'}} \sum_{c \in C} x_{\ell,c} \leq 1 \quad \forall s \in S, t \in T \quad (9.5)
\]

The constraints say that any arc during any interval of length \( h \) (the headway time) can accommodate at most one service a non-zero amount of rolling stock.

The second set of constraints deal with the riding direction on the disrupted link.
\[
\sum_{c \in C} x_{\ell,c} \leq \alpha_t \quad \forall a \in \text{INO}, t \in IT, \ell \in L_a : dt_{ds_a}(\ell) \leq t \leq at_{as_a}(\ell) \quad (9.6)
\]

\[
\sum_{c \in C} x_{\ell,c} \leq \beta_t \quad \forall a \in \text{ISO}, t \in IT, \ell \in L_a : dt_{ds_a}(\ell) \leq t \leq at_{as_a}(\ell) \quad (9.7)
\]

\[
\alpha_t + \beta_t \leq a_{in_t} \quad \forall t \in IT \quad (9.8)
\]

Constraints (9.6) and (9.7) make sure that services can use the disrupted arc only at those time periods when the arc is open for their riding direction. \(at_{as_a}(\ell)\) is the arrival time of service \(\ell\) to the arrival station \(as_a\) of arc \(a\) and \(dt_{ds_a}(\ell)\) is the departure time of service \(\ell\) from the departure station \(ds_a\) of arc \(a\). Constraints (9.8) express the infrastructure limitation to one direction at a time (\(a_{in_t} = 1\)) or to no traffic at all (\(a_{in_t} = 0\)). The value \(a_{in_t} = 2\) indicates no infrastructure limitation in time period \(t\).

### 9.4.5 Passengers Constraints

As mentioned in Section 9.3.3, the passenger demand of an arc is defined for time intervals, denoted here by \(TD_a\). The following constraint links the allocated capacity to the number of passengers \(p_{f_{a,\tau}}\).

\[
\sum_{\ell \in L_a} \sum_{t \in IT : \ell \in L_a : dt_{ds_a}(\ell) \leq t \leq at_{as_a}(\ell)} \sum_{c \in C} \text{cap}_c x_{\ell,c} \geq p_{f_{a,\tau}} - dp_{a,\tau} \quad \forall a \in A, \tau \in TD_a \quad (9.9)
\]

The constraints say that for each arc \(a \in A\) and each time interval \(\tau \in TD_a\), the combined capacity of the trains on the arc during the time interval is enough to accommodate the passenger demand minus the denied passengers. Here \(p_{f_{a,\tau}}\) is the passenger demand in each arc \(a, \tau\) obtained from the multinomial logit model, while \(\text{cap}_c\) is the capacity in composition \(c\).

### 9.4.6 Rolling Stock Constraints

\[
\sum_{c \in C} x_{\ell,c} + y_{\ell} = 1 \quad \forall \ell \in L^p \quad (9.10)
\]

\[
\sum_{c \in C} x_{\ell,c} \leq 1 \quad \forall \ell \in L^c \quad (9.11)
\]
Constraints (9.10) state that each planned service is either canceled or it get exactly one composition. Constraints (9.11) express that emergency services get at most one composition.

\[
\begin{align*}
\sum_{c' \in C} cn_{c,c'} \cdot \delta_{c,c'} + \sum_{s' \in SC} em_{c,s,s'-t_{s,t}} = 1 \\
\sum_{c' \in C} cn_{c,c'} \cdot \epsilon_{c,c'} + \sum_{s' \in SC} em_{c,c',s,s'-t_{s,t}} = 1 \\
\sum_{c' \in C} \delta_{c,c'} + \sum_{s' \in SC} em_{c,c',s,s'-t_{s,t}} = 1
\end{align*}
\] (9.12)

Composition conservation constraints (9.12) ensure the train units’ flow balance. The schedule is given by \(\alpha_{t,s,t}\), which takes the value 1, -1 or 0, if train service \(\ell\) arrives, leaves or stays in station \(s\) at period \(t\), respectively. \(cn_{c,c'}\) is the number of compositions \(c\) needed to obtain a composition \(c'\) in case of aggregation (the number of compositions \(c'\) obtained from composition \(c\) in case of disaggregation). \(et_{s,s'}\) is the travel time between stations \(s\) and \(s'\). Finally, shunting times are given by \(r_s, e_s, d_s\). They are the rotation time duration in station \(s\), the needed time for train unit aggregation in station \(s\) and the needed time for train unit disaggregation in station \(s\), respectively.

\[
\sum_{\ell \in L} x_{\ell,c} + \sum_{c' \in C} \epsilon_{c,s,t-e_s} + \sum_{c' \in C} \delta_{c',c,s,t-d_s} = 1
\] (9.13)
Rotation and departure constraints (9.13) ensure that a rotation is performed before each train service departure. Fleet capacity constraints (9.14) ensure that the number of train units used at time $t \in T$ is limited by the size of the fleet. Note that these constraints count the running trains and those ones in depot stations. Depot capacity constraints (9.15) ensure that the total capacity is not exceeded. The number of parked and shunting train units must be lower than the available capacity. $tu_c$ is the number of train units in composition $c$. Each train service $\ell$ time duration is given by $\beta_{\ell,t}$, which takes value 1, if train service $\ell$ is rolling at period $t$; 0, otherwise. Similarly, $\xi_{s,s',v',t}$ gives information about performance time of an empty train service, which departed from $s$ during $t'$ and is going to $s'$. For composition changes and rotations there are similar parameters with the information regarding the operation time duration, $\mu_{s,v,t}$ and $\gamma_{s,v,t}$, respectively.

\[\sum_{c \in C} tu_c y_{c,s,t} + \sum_{\ell \in L} tu_{c,\ell} x_{\ell,c,t} + \sum_{t' \in T} \sum_{c \in C} \mu_{s,v,t} \left( tu_c \epsilon_{s,v,t}^{c,c'} + tu_c \delta \epsilon_{s,v,t}^{c,c'} \right) + \sum_{t' \in T} \sum_{c \in C} tu_c \cdot \gamma_{s,s',v',t} \beta_{s,v,t} \leq \text{cap}_{s,t} \forall s \in SC, t \in T (9.15)\]

Constraints (9.16) count the number of composition changes in every depot station and during each time period. Note that for all the composition changes that are not
physically possible (i.e., due to composition incompatibility), the variables $\epsilon_{c,c}^{s,t}$, $\delta \epsilon_{c,c}^{s,t}$ are fixed to zero value. Constraints (9.17)–(9.18) denote that the inventory during the initial and final period must be equal to the scheduled one during those time periods, respectively. $t_i, t_f$ are the initial and final time periods in the planning period, respectively. $y_{t_s, 0}^c, y_{t_f}^c$ are the train inventory at depot stations at the initial time period and the final time period in the planning period, respectively.

9.5 Computational Experiments

Our experiments are based on realistic cases drawn from RENFE’s regional network in Madrid for 2008 (Figure 4.1).

We used for our tests personal computer with an Intel Core 2 Quad CPU at 2.83 GHz and 8 GB of RAM, running under Windows 7 64-Bit, and we implemented the models in GAMS/Cplex 12.1.

The Disrupted Network

Our case study features a disruption where one of the two tracks between two stations is blocked: trains in different directions must share the remaining track. Also, some trains that were supposed to pass may turn back instead of entering the disrupted segment. The disruption starts at 8:00 a.m., and it lasts 120 minutes.

The disrupted segment is only used by trains belonging to the C5 line. The alternative paths for passengers of the C5 line include trips on the lines C3, C41, C42 and C5 (run by RENFE) as well as trips on the Metro network (run by another operator). Therefore we restrict the network to these lines only. The restricted network, depicted in Figure 9.1 features 46 station, and about 12,000 trips in 760 timetable services. About 530,000 passengers use the restricted network, 47,000 of which are directly affected by the disruption.

The frequency on the C5 line is rather high: there is a train service every 3 minutes in the rush hours and every 10 minutes in the valley hours. Lines C3, C41 and C42 have a slightly lower frequency: trains in the rush hours run every 6 minutes and every 16 minutes in the valley hours. The considered lines are served with two train units types with a capacity of 588 and 757, respectively. Trains in the Metro system run every 3 minutes and we assume that they have unlimited capacity.
Passenger Demand Split

The disruption has no direct effect on the passengers of lines C3, C41 and C42; they will just stick to their intended path. Passengers of line C5, however, may have multiple traveling options: they can remain in the line C5 waiting for a direct or indirect train; they also can make use of lines C3, C41 and C42 as well as of the Metro network. Depending on their origin and destination, passengers of line C5 may choose any combination of these mentioned alternatives. For example, they may use line C5, then transfer to line C41, and finally come back again to line C5 for the last part of their journey.

In an undisturbed scenario the entire demand is covered by the passenger services. Once the disruption has started, we compute the new passenger demand with the multinomial logit model. Consequently, an anticipated schedule is needed. This anticipated schedule is obtained considering the following: the demand is asymmetric and the track capacity is reduced. The best schedule for passengers would be the one providing proportionate frequency values to the demand in each direction. Then, once the track capacity and the demand to be attended (in an undisturbed scenario) are known, the anticipated timetable is provided to the multinomial logit model.

In our case study, the multinomial logit model gives the following output. In total, 26.2% of the demand in line C5 choose to stay in line C5; 44.3% will go through a combination of lines C5 and lines C3, C41 and C42; finally, 29.5% of the demand will go for a combination of line C5 and the Metro network.
9.5.1 Recovery Solutions

In this section we solve the integrated optimization model for the presented disrupted network. The model’s objective function is a combination of different terms, their relative importance can express different overall managerial goals. Below we consider the same disruption with several different objective weight settings, and we discuss the corresponding optimal solutions.

The optimization model can make timetabling decisions (cancellation of existing services or insertion of an emergency services) as well as rolling stock decisions on the disrupted line C5. In addition, the model may change the rolling stock schedules of the undisrupted lines C3, C41 and C42 in order to adjust the train capacities to the elevated demand figures. We do forbid, though, the cancellation of any of the C3, C41 and C42 services.

Line C5 is independent from lines C3, C41 and C42, they do not share any rolling stock resources. Therefore the optimization model decomposes into two independent subproblems: one for the disrupted line C5, and one for the undisrupted lines C3, C41 and C42. Below we present the solutions.

We solve five different variants of the ITRSRM model. Table 9.1 summarizes our results by letting each column represents one of the five ITRSRM solutions. Each column contains five characteristics of the given solution. Rows TSOC and EMOC give the total operational costs for passenger train services and empty movements, respectively. Row DP gives the number of denied passengers. SC is the number schedule changes which is a measure to estimate how easy is it for the operator to implement the recovery plan; SC accounts for rolling stock changes, empty movements changes and cancellations. Finally, row ST gives the solution time in seconds.

We want to emphasize that the optimization model considers arc-based demand figures. Therefore, the number of denied passengers will also be arc based. However, this is a heuristic approach since a denied passenger would still be counted by the model on the following arc. Row DP-est is the number of denied passengers as estimated by the optimization model, while row DP gives the exact number of denied passengers and is calculated in a post-processing step.

The solution $P\&O$ arises by minimizing the combination of passenger and operator costs (i.e., all the terms in the objective function).

The solution $P\&O-RS$ is obtained by minimizing the combination of passenger and operator costs subject to the additional constraint that all non-cancelled services must keep their originally planned rolling stock composition. That is, the model can only
cancel trains or add emergency services.

The solution $P\&O-RS-EM$ is similar to $P\&O-RS$; the additional constraint is that non-canceled empty trains must be unchanged compared to the undisrupted schedule.

The solution $Operator$ is obtained by minimizing the operator’s costs only.

Finally, the solution $Pax$ is obtained by minimizing the number of denied passengers (DP) as a sole objective. Rolling stock related costs are not taken into account and we assume an unlimited fleet size. In fact, the solution uses more than 300 empty movements. The objective value of $Pax$ is a reference lower bound for the passenger service quality of the other, more realistic, solutions.

We first notice that the lowest possible DP is very well approached whenever the passenger costs are part of the objective (solutions $P\&O$, $P\&O-RS$ and $P\&O-RS-EM$). The service quality deteriorates slightly as we impose more and more restriction on the schedules. Solution $Operator$, on the other hand, results in almost 15,000 denied passengers, 8 times more than what is be achieved in another solution. We will elaborate on the gap between DP and DP-est below in Section 9.5.2.

The operational costs (TSOC and EMOC) do not show much variation. As one may expect, $Operator$ is the best, but $P\&O$, $P\&O-RS$ and $P\&O-RS-EM$ are reasonably close to it. Solution $Pax$ cannot be compared to the other ones as it uses unlimited rolling stock and a huge number of empty movements.

The ease of the recovery process is measured by the number of schedule changes (SC). The results are conform with our intuition: solutions $P\&O-RS-EM$, $P\&O-RS$ and $P\&O$ have increasingly more freedom to change the schedules, and they indeed use this freedom to reach a better service quality. Based on preliminary discussions with practitioners, the SC values of 25-40 all have a good chance to be implementable in practice. Solution $Operator$ needs an even higher ST in order to improve slightly on the operational costs. Solution $Pax$ has little practical value.

The solution times (ST) range from a few seconds to a few minutes. Therefore the proposed model fits well in the time frame of real life disruption management.

### 9.5.2 Comparison of $P\&O$ and $Operator$.

In this section we have a closer look at two radically different solutions: $P\&O$ and in $Operator$. We omit $P\&O-RS$ and $P\&O-RS-EM$ because their characteristics are very similar to those of $P\&O$.

Figures 9.2 and 9.3 depict the passenger demand as a function of time; Figure 9.2 belongs to $P\&O$, Figure 9.3 belongs to $Operator$. Each chart contains two curves: the
higher curve indicates the sum of all passenger demand figures at a given time, while the lower curve shows the denied passenger demand at that time.

The denied demand in solution *P&O* (Figure 9.2) is indeed a very small fraction of the whole demand. This gives an empirical validation for our passenger modeling approach: in spite of the simple per-arc demand structure, the model treats the overwhelming majority of passengers accurately; in fact, the assigned capacity covers most of the passenger demand. This validation does not hold for the case of *Operator* (Figure 9.3). By disregarding the altered passenger demand pattern, the model chooses insufficient train capacities and thereby heavily overestimates the number of denied passengers. Even the correct DP value is very high, denied passengers constitute a significant part of the total demand.

Figure 9.2 Passenger demand in the solution *P&O*: number of requested trips (higher curve) and number of denied trips (lower curve)

The inclusion of passenger costs guides the model towards a good timetable as illus-
Figure 9.3 Passenger demand in the solution Operator: number of requested trips (higher curve) and number of denied trips (lower curve)

Table 9.2 Frequency values in Line C5 between VIAL and OR

<table>
<thead>
<tr>
<th>Direction</th>
<th>Operator P&amp;O</th>
<th>Undisrupted</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIAL→OR</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>OR→VIAL</td>
<td>22</td>
<td>18</td>
</tr>
</tbody>
</table>

trated by the following observation. The disrupted line segment allows for one train at a time to pass. Table 9.2 shows how many trains per direction are scheduled in the two considered solutions. The undisrupted solution is asymmetric, so is the demand: there are more trains from the suburbs towards the city center during the morning rush hour. Solution P&O maintains the same pattern, even though the number of passing trains is reduced. However, solution Operator behaves differently because the disruption appears to create a relative rolling stock shortage on the suburban side.

Finally, we analyze the price of the recovery plan P&O as compared with the original rolling stock assignment. The results are summarized in Table 9.3. We give more details about the solutions in Cadarso et al. [38].

The recovery schedule has lower operating costs for train services which follows from the fact that some train services are canceled. On the other hand, the recovery schedule has higher empty movements costs because more empty movements are needed in order to match capacity requirements and rolling stock resources. By definition, the original schedule has no schedule changes. Finally, we notice that the recovery schedule needs a higher
Table 9.3 Recovery and original planning

<table>
<thead>
<tr>
<th>Schedule</th>
<th>TSOC</th>
<th>EMOC</th>
<th>#SC</th>
<th>#DP</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovery</td>
<td>166338.06</td>
<td>6984.63</td>
<td>40</td>
<td>5221(1839)</td>
<td>206</td>
</tr>
<tr>
<td>Original</td>
<td>167937.28</td>
<td>5355.87</td>
<td>–</td>
<td>–</td>
<td>122.5</td>
</tr>
</tbody>
</table>

computational time, among others due to the additional complexity of the timetabling decisions.

9.5.3 Iterative use of the two-step approach

The results in Sections 9.5.1-9.5.2 arise from a single use of the two-step approach where the passenger demand is computed for an anticipated fictitious timetable. Recall that the passengers’ path choice depends on the timetable but not on the rolling stock assignment. One may wonder how stable is this demand: does the optimized timetable lead to another travel path choice than what we anticipated?

In order to answer this question, we embed the basic two-step algorithm in an iterative framework. The first iteration is what we have done so far. In each subsequent iteration, we re-compute the passenger path choice based on the last iteration’s timetable. The results in this section are limited to the objective of \( P\&O \).

It turns out in our computational experiments that the passenger demand barely changes after the first iteration. When compared to the first iteration’s passenger path choice, as few as 619 passengers decide to choose a different path in the second iteration; this is very little compared to the 47,000 passengers of the C5 line during the disruption. The 619 passengers are roughly equal to the capacity of just one rolling stock unit. Spread over a 2-hour time period and 23 departure stations, the 619 passengers may very well be less than the daily fluctuation of the passenger numbers.

Figures 9.4, 9.5 and 9.6 give some deeper insight of these numbers by showing the passenger demand spread in time (during the disruption) for the first two iterations. Figures 9.4 and 9.5 plot the demand for two particular alternative routes (namely: the combination of lines C5 and C41/C42, and the combination of line C5 and the Metro network, respectively). Figure 9.6 shows the demand that remains on line C5. Mind that the graphs have different vertical scales.

There are two curves in each of the graphs: The lighter gray curve represents the demand in the first iteration and the black one the demand in the second iteration. The vertical difference between the curves never exceeds the value of 300. As a consequence, the second iteration leads to a timetable and rolling stock schedule that is almost identical.
to the output of the first iteration. The changes of the second iteration’s solution, as compared to the first’s, are as follows.

Figure 9.4 Passenger demand in the first and second iterations for the combination of lines C5 and C41/C42

Figure 9.5 Passenger demand in the first and second iterations for the combination of line C5 and the Metro network

- The two timetables run the same set of services; seven emergency services are shifted by 3 minutes, and one service is shifted by 7 minutes.

- There are no rolling stock changes.

- The second iteration has 1718 denied passengers. That is, 121 additional passengers are able reach their destination within the restricted network; these passengers were denied in the first iteration.
The total travel time for passengers during the disruption is similar in both iterations: in the first iteration it was 742561.5 minutes and in the second one it is 743972.5 minutes. However, the total travel time before the disruption was 623910 minutes. Obviously, the total travel time is increased during the disruption. Regarding the average travel time per passenger we have that it was 13.27 minutes before the disruption, 16.18 minutes in the first iteration and 16.25 in the second iteration (this average travel time is obtained for passengers traveling from 8:00 a.m. to 10:00 a.m.).

The virtually immediate stabilization of the iterative approach is largely due to a successful initial estimation of the train frequencies on the disrupted segment. More complex disruptions and severe rolling stock limitations may result in timetables that cannot match the initially estimated train frequencies. In such cases the iterative framework is likely to be essential in order to obtain a realistic passenger path choice.

### 9.6 Summary

The proposed algorithmic framework allows us to find solutions for rather different managerial goals in a matter of minutes. The solution $P&O$ turns out to provide a particularly good balance between the optimization criteria: the passengers’ costs and the operator’s costs are simultaneously brought near to their respective lowest possible values.

When the passenger costs are part of the objective function, the denied demand turns out to be a very small part of the whole demand. This gives an empirical validation for
our passenger modeling approach in that the model treats the overwhelming majority of passengers accurately.

In order to better capture the dynamic nature of the passengers’ path choice, we implement our basic two-step algorithm in an iterative setting. The iterative framework does not improve the solution quality significantly in our experiments. The first iteration appears to estimate the passenger behavior well enough. Consequently, there are barely changes in the schedule of the second iteration. We do believe, though, that more complex disruptions may benefit from multiple iteration.
Part III

AIRLINE APPLICATIONS
Chapter 10

INTRODUCTION

In this chapter the airline industry is introduced through a brief historical review. Then, the airline planning process is briefly introduced.

10.1 The Airline Industry Evolution

The airline industry provides a service to every country in the world, and plays an integral role in the global economy. During much of the development of the airline industry, its growth was enabled by major technological innovations such as the introduction of jet aircraft for commercial use in the 1950s, followed by the development of wide-body jets in the 1970s. At the same time, airlines were heavily regulated throughout the world, creating an environment in which technological advances and government policy took precedence over profitability and competition. The established laws with the Chicago Convention in 1947 defined a strongly regulated system, with low margin for competition between airlines (Benito [8]). The only traffic with a certain level of liberalization was the non-regular one, that in Europe relied on the Paris Agreement of 1956.

It has only been in the period since the economic deregulation of airlines, beginning within the USA in 1978, that cost efficiency, operating profitability and competitive behavior have become the dominant issues facing airline management. In the European Union the liberalization has been slower due to two different reasons. The first one, the Treaty of Rome of 1957 divided the transportation modes into continental, maritime and air modes. For the continental modes (road, railway and river) the competition was encouraged. However, for the other two modes, the corresponding authority established the operative regimen. Moreover, the European Commission thought that the intervention was mandatory in order to encourage the competition and avoid the main airlines to
become the sector in an oligopoly.

Airline deregulation or, at least, liberalization has now spread far beyond the USA to most of the industrialized world, affecting both domestic air travel within each country and, perhaps more importantly, the continuing evolution of a highly competitive international airline industry.

Under this deregulated scenario, the airline industry has been a leader in the development and application of operations research methods. Airlines have been using operations research techniques since the 1950s (Barnhart and Talluri [5]) but we can point the major applications since the late 1970s, when the United States airline industry was deregulated. The existing airlines had to choose between getting low the fares in order to match the new companies fares, or losing market share. However, some airlines decided to start with operations research for maintaining their competitiveness. The answer was revenue management. By creating multiple fares and allocating seat inventory among the fare classes, existing airlines could match the low-cost carrier fare for a portion of their seat inventory and charge higher fares for high value customers (typically business travelers). Because the new entrants did not have the operations research and information technology expertise needed to design and implement revenue management systems, existing carriers were able to survive deregulation and the challenge of low-cost entrants (Barnhart and Cohn [9]).

Revenue management represents only one of many operations research success stories for airlines. Another is airline schedule planning. Schedule planning involves designing future airline schedules to maximize airline profitability. In designing schedules, the airline answers questions of which origin-destination markets to serve, with what frequency, and over which hubs. Besides, the airline determines the departure time and aircraft type for each of these flights. With thousands of flights per day, hundreds of aircraft of different types, hundreds of airports and multiple hubs, constraints related to gates, airport slots, air traffic control, maintenance and crews, and complex issues involving competition, demand forecasting, pricing, and revenue management, schedule planning poses huge challenges for the schedule planner.

Since the deregulation of US airlines in 1978, the pressure on governments to reduce their involvement in the economics of airline competition has spread to most of the rest of the world. US airline deregulation is perceived as a success by most other countries, as the overall benefits to the vast majority of air travelers have been clearly demonstrated. However, it also had some negative effects. The pressure to cut costs, combined with increased profit volatility, mergers and bankruptcies of several large airlines, led to periodic
job losses and reduced wages.

On a global scale, the airline industry has been in a financial crisis for much of the twenty-first century. The problems that began with the economic downturn at the beginning of 2001 reached almost catastrophic proportions after September 11, 2001. At the same time, airline labor costs and fuel prices had been increasing faster than the general rate of inflation for several years. Moreover, airlines faced deteriorating labor/management relations, aviation infrastructure constraints that led to increasing congestion and flight delays, and dissatisfied customers due to perceptions of poor service. Reductions in flight schedules due to the crisis period alleviated some of the pressure on the aviation infrastructure, resulting in fewer flight delays.

The airline industry continues a dramatic restructuring that involves even more fundamental changes than those experienced following its deregulation in 1978. Yet, three decades after the initial deregulation of US airline markets and the subsequent liberalization of many other markets around the world, the industry remains fragile. Multiple cycles of financial successes and failures have left airlines struggling to find a business model that can ensure sustained profitability. Competitive pressure from low cost carriers, the loss of consumer confidence in the air transportation system’s reliability and operating performance, and the transparency of pricing facilitated by the Internet and online travel distribution channels have all contributed to a precipitous decline in average fares and have had a significant impact on airline revenues.

The recent growth of low-fare air travel options combined with a reduced willingness on the part of business travelers to pay the higher airfares charged by legacy carriers contributed in a major way to the poor financial performance of traditional airlines. While it is true that the legacy airlines around the world encountered a revenue problem with the emergence of low cost carriers as formidable competitive force, it also became clear that they had fundamental operating cost and productivity disadvantages compared to their low-cost challengers. The differences in their cost structure reflected substantial differences in the productivity of both aircraft and employees.

The low cost carrier sector took advantage of the weaknesses of the legacy carriers during their financial crises and restructuring efforts. They were able to expand their networks rapidly and captured significant market share. However, during the same period some of the more established low cost carriers also began to face increasing operating costs, driven by aging fleets and personnel with increasing seniority. Furthermore, they could not escape the impacts of increasing fuel costs.

In 2008, after appearing to recover from its latest cycle of financial struggles with
positive financial results in two consecutive years, the global airline industry entered another turbulent period. With the surge in oil prices since 2006, fuel emerged as the single largest industry expense. Even as fuel prices dropped rapidly in late 2008, the very real threat of a deep economic recession driven by unstable financial markets began to weaken the demand for air travel worldwide.

As a consequence of the liberalization of the air transportation mode, different airlines began entering alliances. Thus, they were able to offer a greater number of routes covering the whole world without operating directly the entire network. These alliances usually design their schedules in order to provide smoother flight connections, combine fidelity programs and share different services. Until now, these alliances have been reversible (Benito [18]). However, due to the last financial crisis and increases in fuel costs mergers are spreading through different airlines, which require the competition authority approval. This approval is not for free and it requires to match some requirements to the merger members (i.e., the abandonment of slots in certain airports).

These mergers are resulting in huge airlines with a lot of different resources to schedule in order to provide quality services to passengers. Therefore, mergers will suppose a challenge to planners who have to face huge scheduling problems.

10.2 The Airline Planning Process

In this section the main airline planning phases are introduced. Because it is impossible to simultaneously represent and solve the entire airline schedule planning problem, the many decisions required in airline schedule planning have historically been compartmentalized and optimized in a sequential manner (Lohatepanont and Barnhart [90]).

The decision making is differentiated by time horizon. We can distinguish the following planning levels:

- Strategic: investment decisions are made for the next multiple years in advance such as network design and location of major facilities like airports.

- Tactical: resource assignment decisions are made while looking forward several months.

- Operational: this phase is characterized by several hours or days of planning. It can be further differentiated into ‘off line’ or preventive planning and ‘on line’ or palliative planning.
Different planning phases are listed and explained sequentially. The most important planning decisions faced by airline managers can be summarized as follows:

- Fleet planning
- Route planning
- Schedule development
  - Frequency planning
  - Timetable development
  - Fleet assignment
  - Aircraft routing
- Crew Scheduling

More tactical decisions related to marketing and distribution are also required closer to flight departure, involving pricing and revenue management.

In the following subsections each of the above mentioned planning problems are shortly introduced.

10.2.1 Fleet Planning

Fleet composition is among the most important long-term strategic decisions for an airline, in terms of both its planning process and, ultimately, its operations. An airline’s fleet is described by the total number of aircraft that an airline operates at any given time, as well as by the specific aircraft types that comprise the total fleet. Each aircraft type has different characteristics related to technical performance, the most important of which determine its capacity to carry payload over a maximum flight distance, or range.

Decisions made by an airline to acquire new aircraft or retire existing aircraft in its fleet have direct impacts on the airline’s overall financial position, operating costs and especially its ability to serve specific routes in a profitable manner. The decision to acquire a new aircraft by an airline represents a huge capital investment with a long-term operational and economic horizon.

The impacts on an airline’s financial position of such an investment include depreciation costs that typically are incurred for years, as well as increases in long-term debt and associated interest expenses. From an operational perspective, the decision to acquire a specific aircraft type can have an even longer impact, as some commercial aircraft
have been operated economically for more than 30 years (i.e., McDonnell Douglas DC-9) (Belobaba \[16\]).

It is therefore somewhat surprising that the decision support tools used to make these very important long-term decisions are not as sophisticated as one would expect (or as sophisticated as some of the tools available to airlines for more tactical decisions like scheduling and revenue management). The highly uncertain nature of conditions years into the future has limited the development and use of detailed optimization models for airline fleet planning. Instead, most airlines rely primarily on (relatively sophisticated) spreadsheet-based financial models to make fleet planning decisions (Belobaba \[16\]).

10.2.2 Route Planning

Given the airline’s choice of aircraft and a fleet plan that determines the availability of aircraft with different capacity and range characteristics, the next step in the airline planning process is to determine the specific routes to be flown. In some cases, the sequence of these decisions is reversed, in that the identification of a profitable route opportunity might require the acquisition of a new aircraft type not currently in the airline’s fleet (Belobaba \[16\]).

Economic considerations and expected profitability drive route evaluations for most airlines. Route profitability estimates require demand and revenue forecasts for the period under consideration. In large airline networks, traffic flow support from connecting flights can be critical for route profitability. With the evolution of connecting hub networks around the world, very few flights operated by network airlines on a route carry only local Origin-Destination (O-D) passengers.

Hub-and-spoke network structures allow airlines to serve many O-D markets with fewer flight departures, requiring fewer aircraft generating fewer available seat per kilometer at lower total operating costs than in a complete point-to-point route network. The hub airline’s ability to consolidate traffic from many different O-D markets on each flight leg into and out of the hub allows it to provide connecting service even to low demand O-D markets that cannot otherwise support non-stop flights.

Large hub networks result in market share advantages that translate into increased revenue for the airline (in addition to the reduced operating costs described above). With the potential for the airline to offer greater (connecting) departure frequency in many O-D markets, more convenient schedules can lead to higher market shares against competitors. At the same time, schedule dominance of local markets into and out of the hub may lead to pricing and revenue advantages for the hub airline in those markets.
The route selection decision is both strategic and tactical. It is an essential component of an integrated network strategy for the airline, which must decide whether to focus on short-haul or long-haul services, domestic or international operations. At the same time, the characteristics of the selected routes will affect the types of products the airline offers to travelers. For example, an international route network will likely lead to a decision that business- and first-class products should be offered in order to be competitive.

The distance of the selected routes will also affect the airline’s cost structure, as longer routes will likely be flown with bigger aircraft that have lower unit costs per seat and per available seat per kilometer. Although fleet planning was introduced first, it is important to recognize that the aircraft performance requirements for specific routes can and do provide a feedback loop to fleet planning decisions.

Route planning can also be a much shorter-term tactical process, as unexpected route opportunities often present themselves to the airline with changes to the market environment. For example, the bankruptcy of another airline, a withdrawal from a route by a competitor, or a newly negotiated bilateral agreement with another country can lead to new route opportunities that must be acted upon within months or even weeks.

Economic considerations dominate route evaluation, especially for airlines operating in competitive environments with a profit maximization objective. The most important inputs to any route evaluation are forecasts of potential passenger demand (as well as expected revenues) for the proposed route. Once a forecast of the total O-D market demand (per period) for the route in question has been generated, an equally important step is the estimation of the market share that the airline can expect of this total demand. The airline’s own market share of total forecast demand will depend on its frequency share in the market, the path quality of its planned services (non-stop versus connecting flights), as well as its planned departure times. To the extent that the competitive marketplace will allow differences to be maintained, relative prices and service quality can also have a significant impact on expected market share of total demand.

10.2.3 Airline Schedule Development

The airline schedule development is composed of several planning phases: the schedule design, which is composed of frequency and timetable planning, the fleet assignment and the aircraft routing.
Schedule Design

Schedule design is a critical stage of in airline’s planning process. When a flight schedule is given, a major proportion of costs and revenues are fixed. All the subsequent planning stages have to optimize the use of resources in the space restricted by the schedule. Therefore, optimization of the flight schedule is of great importance to an airline. In most airlines, the flight schedule is drafted several months before it is put into execution. When the first draft comes out, it is studied by the various departments involved in the work of fleet assignment, crew scheduling, aircraft maintenance, and other resource allocation processes. After the draft’s feasibility and economics are evaluated and changes are recommended, it is sent back to the flight planning department for revision. Normally, a flight schedule goes through an iterative procedure of this kind many times before reaching its final.

Given the estimated demand for travel, an airline wishes to determine the flight schedule which maximizes its profit while taking into account the satisfaction of its customers. In this system, two components interact: the aircraft flow in the physical network and the passengers using flights to travel.

The objective of schedule design is to develop a schedule, defining an origin, a destination, a departure time, and an arrival time for each service to accommodate passenger demands given available resources. A market is defined by an origin and destination pair. For example, Madrid-Barcelona is a market, and Barcelona-Madrid is another distinct market, referred to as an opposite market. In schedule design, we are interested in its unconstrained market demand, that is, the maximum demand the airline is able to capture. Unconstrained market demand is allocated to itineraries in each market. The demand is unconstrained because the quantity of interest is measured without taking into account capacity restrictions.

Schedule design is typically composed of two sequential steps: frequency and timetable planning.

Frequency Planning. Increases in the frequency of departures on a route improve the convenience of air travel for passengers. The airline can benefit from higher traffic and revenues associated with this increased frequency, in terms of both more passengers willing to fly given the more convenient air travel services and increased market share at the expense of its competitors.

Peak departure times are most attractive to a large proportion of travelers in many markets. More frequent flight departures further reduce the wait time between flights, reducing travel inconvenience for more passengers. Recall that frequency is much more
important in short-haul markets than for long-haul routes where actual flight time dominates wait time. In competitive markets, airline frequency share is most important for capturing time-sensitive business travelers (Belobaba [16]).

Although the determination of the optimal frequency of service on a route does involve elements of competition, in most cases an analysis of demand forecasts and expected market share can be used to establish a baseline frequency of flights on a route that can be operated profitably by the airline. Based on the outputs of the route evaluation process, the airline should already have estimates of the total demand between the origin and destination served by the route under consideration. The airline’s expected market share of this total demand is then determined by its frequency share and specific timetable of flight departures, relative to its competitors on the same route.

Frequency planning is introduced here as being separate from the fleet assignment for each flight on the route; however, it should be clear that the two decisions are interrelated. The airline’s supply decision for a route in fact consists of two simultaneous choices: the number of departures per day and the number of seats to be offered on each departure.

Airline Competition and Market Share. In general, if there is more than one airline in the market, one carrier’s market share and revenue depend not only on its own service but also on the services provided by all other airlines in the market.

Airlines compete for passengers and market share based on the following factors (Belobaba [15]):

1. Frequency of flights and timetable on each route served.
2. Fares, relative to other airlines.
3. Quality of service.

Passengers choose the combination of flight schedules, prices and product quality that minimizes their total disutility. Each passenger would like to have the best service on a flight that departs at the most convenient time, for the lowest price. However, passengers are seldom able to find the perfect itinerary and the highest service quality for the lowest possible price, so they must trade off these factors to minimize their disutility, subject to budget constraints.

The market share of an airline is defined as the proportion of total market demand that is captured by the airline. With all else being equal, airline market shares will approximately equal their frequency shares (Belobaba [15]). The assumption of all else being equal requires that the price and service quality differences (apart from frequency of service) among competing carriers are negligible.
There is widespread acceptance in the airline industry of an S-curve relationship between airline market share and frequency share. The S-curve describes how an airline’s market share grows non-linearly with the frequency share in that market. As Button and Drexler [29] note, it is difficult to document the origins and evolution of this model from the published literature. Early theoretical development and empirical evidence that higher-frequency shares are associated with disproportionately higher market shares was provided in the 1970s, before deregulation (Simpson [113]; Taneja [118]). After deregulation, there exist references to the S-curve (Kahn [81], Baseler [13], Vaze and Barnhart [122]).

**Timetable Planning.** The frequencies in the spatial network are taken as parameter in the timetable development. Establishing a timetable of flight departures requires a trade-off between maximization of aircraft utilization (block hours per day) and schedule convenience for the passengers. The timetable must incorporate minimum “turnaround” times required at each airport to deplane and enplane passengers, refuel and clean aircraft (Belobaba [16]).

Here, the problem of determining the schedule for the flights which have been chosen by solving the frequency planning is considered. We suppose that the schedule must be periodic with period $T$. The problem is to assign a departure time to each flight. The main constraints to be considered in solving the problem are the following (Cadarso [31]):

1. the demand of transportation for passengers must be satisfied,
2. at any point in time, the airport and airway capacities cannot be exceeded,
3. each flight cannot carry more than a given number of passengers,
4. the time interval between the departure times of consecutive flight legs serving the same origin-destination is bounded.

In addition to the constraints described, numerous additional factors constrain an airline’s timetable development. Connecting hub networks that operate on a fixed bank basis require that flights arrive from spoke cities within a prescribed time range, to facilitate passenger connections. This requirement leaves relatively little flexibility for scheduling departures from the spoke city to the hub, if the flights are expected to provide service to both local and connecting passengers. Time zone differences also limit feasible departure times, especially on long-haul routes. Moreover, crew scheduling and routine maintenance requirements also have substantial impacts on timetable development.
Fleet Assignment

The fleet assignment problem is to determine the type of aircraft to be flown on each flight leg departure, given a planned network of routes and specified timetable of flights. It has a tremendous impact on airline’s profits, as it directly affects flight operating costs and passenger revenues.

The factors that influence schedulers when assigning fleet types to various flights are: passenger demand, seating capacity, operating costs, and availability of maintenance at arrival and departure stations. Aircraft are routed after the fleets are assigned to ensure the solution is operational. A good flight schedule should also provide sufficient flexibility to enable efficient crew scheduling to be done. On the other hand, flight schedules are often revised to facilitate feasible or more effective flight assignments.

One important requirement of the fleet assignment is that the aircraft must circulate in the network of flights. These so-called balance constraints are enforced by using time lines to model the activities of each fleet type. In the time line model, there is a network built on the flight schedule for every fleet type. The components of the network for each fleet type are as follows: each flight’s arrival corresponds to a node at the arrival station and at the ready time, the time after which the aircraft can start to fly after the previous flight, and each flight’s departure also corresponds to a node at the departure station and at the departure time. The ready times may differ among different fleet types due to the fleets’ individual physical conditions. Connection each flight’s departure and arrival nodes is the flight arc, and between each node and its next adjacent node on the time line at the same station is the ground arc. There is a ground arc for each station connecting the last node to the first node for a station to complete a daily operational cycle.

Even with fleet assignment optimization models, it is important to recognize that it is not possible to achieve a perfect match of seats with demand on each flight in the timetable, given the need to balance the flow of aircraft over the entire airline network (Barnhart [11]).

Fleet assignment optimization, which has been applied widely in practice, is attributed with generating solutions that lead to significant improvements in operating profit. Examples include the following: Abara [1], Hane et al. [70], Rushmeier and Kontogiorgis [107], and Subramanian et al. [116].

Aircraft Routing

The solution obtained from the fleet assignment identifies the flow of the fleet through the network. However, it does not identify which specific aircraft from that fleet is assigned to
each flight leg. The output of the fleet assignment problem is used as input into the aircraft routing problem. Given an assignment of the fleet to flights, the airline must determine a sequence of flights, or routes, to be flown by individual aircraft such that assigned flights are included in exactly one route, each aircraft visits maintenance stations at regular intervals, and there is always an aircraft available for a flight’s departure. Aircraft routing is the process of assigning each individual aircraft within each fleet to flight legs (Desaulniers et al. [58]). The primary objective of the aircraft maintenance routing problem is to find a feasible solution, one ensuring sufficient maintenance opportunities for each aircraft. Historically, another objective has been to determine routings that maximize through revenue. The aircraft routing is also referred to as aircraft rotation, aircraft assignment or tail assignment. Aircraft are normally distinguished by their tail registration numbers. A tail number is a unique serial number assigned to each aircraft for each airline in each country.

10.2.4 Crew Scheduling

The goal of crew scheduling is to identify cost-minimizing crew schedules that provide the necessary crews for each flight, while satisfying the myriad of constraints imposed by government and labor work rules. A survey of optimization approaches for crew scheduling is provided in Barnhart et al. [8].

Each flight has to be performed by one crew. In fact, each crew performs a roster, defined as a sequence of flights whose operational cost and feasibility depend on several rules laid down by union contracts and company regulations. The problem consists of finding a set of rosters covering every flight of the given planning horizon, so as to satisfy all the operational constraints with minimum cost.

Usually, the crew planning problem is approached in two phases, according to the following scheme:

1. Crew Scheduling: the short-term schedule of the crews is considered, and a convenient set of duties (pairings) covering all the flights is constructed. Each duty represents a sequence of flights to be covered by a single crew within a given planning horizon.

2. Crew Rostering: the duties selected in the crew scheduling phase are sequenced to obtain the final rosters.
10.2.5 Integrated Airline Planning

Although practical, the sequential nature of aircraft and crew schedule optimization leads to suboptimal plans, with potentially significant economic losses. Improved plans can be generated building and solving integrated models of some (and eventually all) of these schedule design, fleet assignment, routing and crew scheduling sub-problems (Barnhart [11]). Another area of integration is to expand schedule planning models to include pricing, revenue management decisions and competition. It should be clear that a joint decision approach can have substantial benefits for the airline. For example, better feedback from pricing and revenue management systems can affect the optimal choice of schedule and aircraft capacity. At the same time, improved decisions related to schedule and capacity can reduce the need for excessive discounting and fare wars.

Joint optimization and planning is a big challenge, both theoretically and practically, for the following reasons. Despite having an enormous amount of detailed booking, revenue and operations data, few airlines have “corporate databases” that provide a single source of consistent and detailed demand/cost data, as required for joint decisions (Barnhart [11]). Furthermore, a great deal of research is still required to identify and calibrate models that can capture market dynamics and competitive behaviors that are necessary for joint optimization of airline planning decisions. Finally, there are obstacles within many airlines in terms of organizational coordination and a willingness to accept a large-scale decision tool that claims to solve several airline planning problems simultaneously. In fact, it might never be possible to integrate all the subtleties of airline planning decisions into a single optimization tool.

10.3 Airline Planning Problems at IBERIA

IBERIA is the main Spanish airline. Based in Madrid, it operates an international network of services from its main base of Madrid-Barajas Airport. In addition to transporting passengers and freight, Iberia Group carries out related activities, such as aircraft maintenance and handling in airports.

Some of the above presented problems focused on passenger transportation will be addressed during this dissertation. All of them will focus on the national network in Spain, which is fully described here (Figure 10.1). All of our computational experience is for realistic cases using data of 2010.
10.3.1 The Air Network

The air network is formed by the airports and all the feasible airway alternatives linking them. The airports are characterized by the operations that can be performed at those airports. They are characterized by available capacity (i.e., slots) for landing and take-offs determined by airports’ operators. The airways are the links between the airports. Each airway is characterized by an origin airport and a destination airport. When an airway is flown it will be called as flight leg; a flight leg is defined by a departure airport-time period and arrival airport-time period. An airport-time period is defined as a combination of an airport and a specific time period at that airport. Each flight leg must be assigned to exactly one fleet type. There are different fleet types in this network. Each fleet type is mainly characterized by its seating capacity and cruise speed.

We have studied a simplified version of IBERIA’s air network: the national network. The air network is a pure hub-and-spoke network with 23 different airports. The hub is located in Madrid. There is not any direct flight bypassing the hub airport. There are 44 possible flights and 104 passenger paths within the network. There are three different fleet types available for this study case: A-319 fleet with 141 seats per aircraft, A-320 fleet with 171 seats per aircraft and A-321 fleet with 200 seats per aircraft.

Figure 10.1 IBERIA’s national network
10.3.2 Passenger Demand

The unconstrained demand is characterized by the combination of origin airport, destination airport and the desired departure time. We will refer to this combination as a market. Consequently, for each market the departure airport-time period is known. The unconstrained total demand of passengers in each market is assumed to be fixed and known. For our study an estimate of such demand is provided by the airline.

In each market, passengers can choose any of the corresponding itineraries. Each itinerary is defined by a set of flight legs that connect the origin and destination airports. It can be composed of one flight leg or it can consist of more than one leg including intermediate stops at different airports.

10.3.3 Part III Outline

In the following chapters some airline optimization problems will be addressed. All of them have been developed and tested in the network presented in 10.3. The presented problems are:

- Robust Passenger Oriented Timetable and Fleet Assignment Integration in Airline Planning and,

Chapter 11

ROBUST PASSENGER ORIENTED TIMETABLE AND FLEET ASSIGNMENT INTEGRATION IN AIRLINE PLANNING

This chapter looks at the airline-scheduling problem and develops an integrated approach that optimizes schedule design, fleet assignment and passenger use so as to reduce costs and create fewer incompatibilities between decisions. Robust itineraries are created to ameliorate misconnected passengers. The analytical work is supported with a case study involving the Spanish airline, Iberia. Our approach shows that the number of misconnected passengers can be reduced when robust planning is applied.

11.1 Introduction

The schedule is an airline’s primary product, having the most influence with price on a passenger’s choice of an airline. Once an airline decides on a schedule, several related problems have to be solved before the schedule is feasible. Consequently, the airline schedule planning problem is defined as the sequence of decisions that need to be made to make a flight schedule operational. These decisions are made in different time horizons. Given the high level of competition in the airline industry, effective decision tools are crucial to the profitability of an airline. This is the motivation for this paper in which
we focus on the integration of the decision making process at tactical level where a time horizon of several months is available for planning. Our goal is to achieve simultaneous rather than sequential solution; a simultaneous solution will generate more economical solutions and create fewer incompatibilities between the decisions. Moreover, with the integration of the different planning process phases a greater robustness degree may be achieved, obtaining smoother solutions, which in case of disruption may be recovered in an easier way.

Legacy airlines usually use hub and spoke networks. Arithmetically adding a flight in a hub and spoke network creates geometric market additions via flight connections. Consequently, airlines are able to attend more markets than with other types of networks. Hub and spoke networks also permit to get lower operating costs because some infrastructures such as resources needed for maintenance may be concentrated in hubs. In hub and spoke networks, connecting passengers are very common. In order to fly from one spoke to a different one, a connection must be performed in the network hub. The time needed to accomplish these connections is established but whether this is enough time will vary due to stochastic events (e.g. weather). Thus, robustness will be introduced through passengers’ itineraries, providing them enough connection time. However, as connection time increases, passengers’ dissatisfaction and low resource utilization may also increase. Therefore, robustness is introduced avoiding misconnected passengers but accounting for passengers’ costs and fleet utilization.

This robustness criterion may have different impact over different aspects. First of all, increasing connection time means that passengers will have to perform longer connections, increasing their dissatisfaction due to the longer connection time. However, the probability of being misconnected will be ameliorated. Another issue might be fleet utilization. Designing the timetable providing robust itineraries without accounting for the fleet will surely provide a lower utilization of it. Nevertheless, this hassle may be controlled by facing the problem in an integrated way, accounting for fleet resources as we will propose later in the paper.

The tactical problem presented in this chapter consists of determining the schedule design and the fleet assignment simultaneously accounting for passenger itinerary demand. Therefore, we will focus on passengers in order to avoid misconnections providing robust itineraries with enough connection time.

The rest of the chapter is as follows. Section 11.2 gives an overview of related literature with the proposed problem. In section 11.3 the problem is described. The mathematical formulation of the problem is shown in section 11.4. In section 11.5 computational
experiments are presented.

11.2 Literature Review

At tactical level, Armacost et al. [4] describe a new approach for solving the express shipment service network design problem. They transform conventional formulations to a new formulation using what they term composite variables. The formulation relies on two key ideas: first, they capture aircraft routes with a single variable, and second, package flows are implicitly built into the new variables, the composite variables. Yan and Tseng [128] develop a model and a solution algorithm to help carriers simultaneously solve for better fleet routes and appropriate timetables. They state that traditional approaches, which employ draft timetable as an essential medium, reveal an incapability of directly and systematically managing the interrelation between supply and demand. Their proposed model is formulated as an integer multiple commodity network flow problem. Lan et al. [85] consider passengers who miss their flight legs due to insufficient connection time. They develop a new approach to minimize passenger misconnections by re-timing the departure times of flight legs within a small time window. They present computational results using data from a major U.S. airline and showing that disconnected passengers can be reduced without significantly increasing operational costs.

Dynamic scheduling is a relatively new approach to flight scheduling. At tactical level demand forecasts are not enough accurate and schedules must be changed as more information reveals. Consequently, dynamic scheduling has a rescheduling component. Jiang and Barnhart [79] propose a dynamic scheduling approach that reoptimizes elements of the flight schedule during the passenger booking process. They redesign the flight schedule at regular intervals, using information from both revealed booking data and improved forecasts. Jiang and Barnhart [80] develop robust schedule design models and algorithms to generate schedules that facilitate the application of dynamic scheduling. Such schedules are robust in the sense that the schedules produced can more easily be manipulated in response to demand variability when embedded in a dynamic scheduling environment.

Lately, researchers have focused on determining incremental changes to flight schedules, producing a new schedule by applying a limited number of changes to the existing schedule. Lohatepanont and Barnhart [90], in their incremental optimization approach select flight legs to include in the flight schedule and simultaneously optimize aircraft assignments to these flight legs. Garcia [66] extends the previous model and proposes a combination between it and a decision time window model. Kim and Barnhart [82]
consider the problem of designing the flight schedule for a charter airline. Exploiting the network structure of the problem, they develop exact and approximate models and solutions, and compare their results using data provided by an airline. Espinoza et al. present an integer multicommodity network flow model with side constraints for on-demand air transportation services.

At tactical level, Cadarso and Marín propose a multiobjective integrated robust approach for the schedule design phase, deciding jointly flight frequencies and timetable. The objectives are passengers’ satisfaction and operator costs. They try to fix the timetable ensuring that enough time is available to perform passengers’ connections, making the system robust avoiding misconnected passengers. Cadarso and Marín extend the previous model assigning fleet types to flight legs in a minimum cost multicommodity network. The authors do not use a draft timetable but it is built from scratch. Here, fleet utilization and some schedule issues (e.g. flights the airline wants to schedule in any case) are not addressed.

11.3 Problem Description

In this chapter the schedule design and fleet assignment problems are treated in an integrated way in order to account for the interrelation between supply and demand. Frequencies and departure times must be determined for every itinerary attending each market. Moreover, fleet types must be assigned to every flight leg. Given the estimated demand for travel, an airline wishes to determine the flight schedule which maximizes its profit while taking into account the satisfaction of its customers. In this system, two agents interact: the aircraft flow in the physical network (supply), and the passengers using flight legs to travel (passenger demand).

11.3.1 Supply

The network is formed by the airports and all the feasible airway or sections alternatives linking them. The airports are defined by the operations that can be performed within them. They are characterized by available capacity (i.e., slots) for landing and taking off determined by airports’ operators.

The sections are the links between the airports. Each section is characterized by an origin airport and a destination airport. When a section is flown it will be called as flight leg; a flight leg is defined by an origin, destination and a departure time, that is, a flight leg is defined by the pair \((s, t)\), where \(s\) is an element of the sections’ set, \(S\), and \(t\) in
\( \{0, 1, ..., T - 1\} \) is the departure time from the origin of \( S \). The set of all possible flight legs is \( F = S \times \{0, 1, ..., T - 1\} \). Therefore, we will consider that the time is discretized by partitioning the planning period, \( T \), into intervals of equal length with starting points \( 0, 1, ..., T - 1 \). The intervals’ length will be taken as the time unit. The time periods’ size may differ depending on the study case; in the forthcoming case they will be 15 minute time periods.

Each flight must be assigned to a fleet type. There are different fleet types in this problem. Each fleet type is mainly characterized by its seating capacity and cruise speed. That is, flight time may depend on the assigned fleet type to the flight. In a tactical problem the number of airplanes to be used is affected by uncertainty and it may vary. However, it makes no sense to plan to use a high number of aircraft if they have a low utilization in the proposed schedule; the low utilization is bad for many reasons such as the need of many crews and the difficulty in amortizing the fleet. Consequently, a minimum average block hour utilization is imposed for every fleet type.

A complete new schedule is not usually welcome in an airline. It has some obligations: some governments give grants in order to maintain a minimum level of service in some markets; the airline wants to schedule a determined number of flights in some markets in order to maintain an image compared to some other competitors; some markets are slot constrained and if the airline does not operate flights it loses slots rights; etc. Consequently, some information about some determined flights is given in order to match all the previous requirements.

### 11.3.2 Passenger Demand

The unconstrained demand is characterized by the origin airport, \( o \), destination airport, \( d \), and the desired departure time \( t \). Each set \( (o, d, t) \) is mentioned as the market \( w \). For each market \( w \) the demand of passengers \( d_w \) is assumed fixed and known datum (given by the airline). Although the market has a desired departure time, it is not a fixed value, passengers will accept without any additional cost a departure time from a set of compatible departure times in each market. Consequently, the airline supposes that passengers have a feasible set of departure times. This set of departure times will depend on the market among others. For example, for markets with long trips this set will be bigger than for markets with shorter trips.

For each market, passengers are considered in all possible itineraries \( i \in I_w \). \( I_w \) is the set of itineraries attending market \( w \). Each itinerary is defined by a set of sections that connect different airports and a departure time. It can be composed of one section
or more than one including in this last case intermediate stops at different airports. If connecting time is not enough, passengers may be misconnected. In order to avoid misconnected passengers as much as possible, robust itineraries are introduced. We define a robust itinerary as that one that minimizes misconnected passengers due to lack of time to perform flight connections. Obviously, connecting time will be a trade-off between airline’s available resources and passengers’ perception with waiting time.

Many markets will face the problem that passengers in them will not have any available flight to fly. So they will choose either a compatible but different market or to travel in a different company. We define compatible markets as those ones with identical origin and destination but different departure times. Therefore, some passengers will choose a compatible market (according to a recapture rate) and some other passengers will choose to travel in a different company. The recapture rate is a measure of the probability of accepting the alternative itinerary. The attractiveness of the alternative is based on the time of day of departure, length of trip, and number of connections. This issue is illustrated in Figure 11.1. For the same origin-destination there are three different itineraries within the same company. It may occur that passengers willing to travel between $k_1$ and $k_5$ have no room in the second flight. Consequently, the airline will offer them two different choices. Traveling between $k_1$ and $k_6$ or traveling between $k_7$ and $k_8$. Depending on the rest of the available itineraries within other airlines the recapture rate will be calculated and some of the disrupted passengers will be recaptured in the offered itineraries.

As competition effects are not considered in this problem, every flight leg will probably be crowded because the unconstrained market demand is considered. However, this situation is far from the real one. Due to competing airlines, unconstrained demand will be divided between different flight legs in real life. In order to represent this issue, the capacity offered for each flight leg will not be the entire one but the one obtained by an average load factor. This average load factor will be obtained from the airline data. Besides, competition within the same airline will be avoided by imposing a separation time between flight legs operating the same origin and destination. Consequently, the model will adjust the unconstrained demand to the available capacity in the scheduled flight legs.

### 11.3.3 Robustness

As mentioned before, robustness is introduced through passengers in itineraries with more than one flight, where a connection is mandated. Adding more slack for connection can
be good for connecting passengers, but can result in reduced productivity of the fleet; the challenge then is to determine where to add this slack so as to maximize the benefit to passengers without getting worse the network operation (Lan et al. [85]).

Every connection is characterized by the minimum time required to perform it. This time varies from airport to airport and it can also vary along the day. In this way, in itineraries with more than one flight, every passenger is mandated a minimum connection time per itinerary \((mct_i)\) for flight connections. However, this time will not be always enough to perform the connection due to several reasons such as weather conditions, operational delays, security checks and crowded airports.

We assume that the number of disrupted passengers depends on the available time to perform the flight connection. Once flights’ arrival and departure times are known, the available connection time per itinerary \((ct_i)\) is also known. From airlines’ historical data, disrupted passengers number variability with connection time may be known.

Assigning a statistical distribution to misconnected passengers, the probability of getting misconnected passengers depending on connection time can be calculated. The exponential distribution has been chosen. \(ect_i\) represents the available excess connection time per itinerary \((ect_i = ct_i - mct_i)\). Consequently, given the available excess connection time \((ect_i)\), the probability of having misconnected passengers \((prob_i)\) is:

\[
prob_i = e^{-\lambda_i(ect_i+t_0)}
\]  

where \(\lambda_i\) depends on the itinerary connection characteristics (such as the airport where the connection is performed, time of day and type of flight) and is chosen adjusting the probability distribution to historical data; it is supposed that once the connection characteristics are known, the assigned gates will be probably known due to historical availability. \(t_0\) is a location parameter to fit the distribution to the available data.

In Figure [11.1] an example is presented. There are three different available itineraries between two different airports. Two of them are composed of two different flights, so a flight connection must be preformed. For itineraries composed of more than one flight we have a minimum connection time \(mct\) and connecting times \(ct1\) and \(ct2\), being \(ct2 > ct1\). Consequently, the probability of having misconnected passengers will be lower in the itinerary with connecting time \(ct2\). Therefore, if we do not attend to the presented robustness criterion, passengers will be assigned to the itinerary with the lower connecting time because that itinerary will have a lower negative perception for them (passengers would like to reach their destination as fast as possible). In the robust approach a trade
off between misconnected passengers and fastest arrival time will be found, probably assignning some passengers to the itinerary with the greater connecting time.

Once misconnected passengers are known, they must be removed from the remaining flights of the itinerary, so extra capacity arises in those flights making possible to accommodate other passengers in it in case of disrupted passengers. Therefore, the airline will overbook flights depending on misconnected passengers.

### 11.4 Integrated Robust Airline Scheduling Model

We propose an integrated and robust model for timetable, fleet assignment and passenger demand optimisation. Consequently, a global optimum solution is obtained for those three problems. The proposed model is based on the one presented by Cadarso and Marín [35], where the authors proposed a minimum cost multicommodity flow problem. However, in this paper we present a revenue maximizing problem; the main reason for this is that we need to protect high value passengers. We also include the minimum average block hour utilization in order to avoid the fact of having aircraft with low utilization. Another extension of the problem is the inclusion of a part of the schedule that is given and mandatory in any case.

Passengers transfer possibility is considered, that is, for every passenger itinerary the
possibility of intermediate stops in the flight are taken into account. Itineraries composed of up to two flight legs are considered. We suppose that the schedule will be periodic, that is, the schedule will repeat after the planning period ends. For this purpose, we must take care about airplanes location at the end of the planning period. Its location must be the necessary one to repeat the schedule in the following planning period.

The following notation is introduced to define the Integrated Robust Airline Scheduling Model (IRASM):

- **Sets:**
  - $F$: set of flights indexed by $f$. They are defined by a departure airport, a departure time and a destination airport.
  - $ROD$: set of pairs origin-destination indexed by $od$. These pairs are subject to some regulation due to airline’s strategy.
  - $P$: set of fleet types indexed by $p$.
  - $T$: set of time periods indexed by $t$.
  - $S$: set of sections indexed by $s$. Each section is defined by an origin airport and a destination airport.
  - $W$: set of markets indexed by $w$. We now define the markets by the origin, $o$, destination, $d$, and the desired departure time $(o,d,t)$.
  - $K$: set of nodes indexed by $k$. Defined by an airport and a time period.
  - $I$: set of itineraries indexed by $i$. They are defined by the departure airport and departure time and the destination airport.
  - $F_{od}$: set of flights operating in pair origin-destination $od$.
  - $I_w$: set of itineraries attending market $w$.
  - $I_f$: set of itineraries using flight $f$.
  - $RE_w$: set of markets compatible with market $w$.
  - $CF_w$: set of fights that compete for demand in the same market $w$.
  - $AS_{k,p}$: set of flight legs with $p$ fleet type assigned that arrive in node $k$.
  - $DS_{k,p}$: set of flight legs that depart from node $k$ with $p$ fleet type assigned.
  - $CT$ is the set of the count time, that is, the time period at which the fleet is counted.
- Parameters:
  
  \( b_i^w \) is the average revenue per passenger in itinerary \( i \) from market \( w \).
  
  \( \tau_i \) is an increasing dummy cost for passengers with the connection time. It represents passengers’ negative perception with increasing connecting time.
  
  \( mpc_i \) is the cost per misconnected passenger in itinerary \( i \).
  
  \( rpc_w^w' \) is the cost of recapturing a passenger from market \( w \) to market \( w' \).
  
  \( \alpha_w^w' \) is the recapture rate from market \( w \) to market \( w' \).
  
  \( cBH_p \) is the cost per block hour for each fleet type \( p \).
  
  \( NBH_f^p \) is the number of block hours in each flight \( f \) with a determined fleet type \( p \).
  
  \( c_p^f \) is the operating cost for flight \( f \) with fleet type \( p \).
  
  \( d_w \) is the passenger demand number in market \( w \).
  
  \( alff \) is the average load factor for flight \( f \).
  
  \( q_p \) is the passenger capacity in fleet type \( p \).
  
  \( qa_k \) are airports’ available slots for arrivals allocated by time-of-day; available slots are very scarce in busy airports during rush hours.
  
  \( qd_k \) are airports’ available slots for departures allocated by time-of-day; available slots are very scarce in busy airports during rush hours.
  
  \( \beta_{f,t} \) takes value 1 if flight \( f \) is flying during \( t \); 0, otherwise.
  
  \( k_i \) and \( k_f \) are the initial and final node for each airport, respectively.
  
  \( m_{od} \) is the minimum level of service in pair \( od \).
  
  \( fu_p \) is the minimum average block hour utilization per day and fleet type \( p \).
  
  \( N_p \) is the fleet size of fleet type \( p \).

- The variables of the problem are the following:
  
  \( z_f^p \): =1, if flight \( f \) is assigned with \( p \) fleet type; 0, otherwise.
  
  \( y_p^k \): integer variable. It represents grounded material of type \( p \) at node \( k \).
  
  \( h_i^w \): integer variable. Passengers in itinerary \( i \) and market \( w \).
  
  \( g_w^w' \): integer variable. Number of passengers the airline tries to send from market \( w \) to market \( w' \).
For the sake of clarity we declared all variables to be integral. We note that the nature of the constrains allows us to relax the integrality of $h^w_i$, $g^w_{w'}$ and $y^p_f$.

The IRASM is defined as a multicommodity flow model on a quite large network, and arises as:

11.4.1 Objective Function

$$\text{max } z = \sum_{w \in W} \sum_{i \in I_w} (b^w_i - \tau_i - mpc_i \cdot prob_i) h^w_i - \sum_{w \in W} \sum_{w' \in RE_w} rpc^w_{w'} \cdot \alpha^w_{w'} g^w_{w'} - \sum_{f \in F} \sum_{p \in P} c^p_f z^p_f \tag{11.2}$$

The objective function (11.2) consists of maximizing airline’s profit. The first two terms are related to passengers and the last one to the operator. The objective function protects high value passengers by maximizing the number of attended passengers multiplied by their itinerary fare. It also accounts for the negative perception perceived by passengers with the increasing connecting time. The number of misconnected passengers is given by $prob_i \cdot h^w_i$. Passengers may be also recaptured. However, it has a negative effect because recapturing passengers means that they were not attended when they asked for it. Operating costs are the costs the company incurs due to the operation of flight legs. We calculate these costs as the product of the number of block hours in each flight $f$ with a determined fleet type $p$ by the cost per block hour for each fleet type. We compute these costs in (11.3).

$$c^p_f = cBH_p NBH_f^p \tag{11.3}$$

The objective function is subject to the following groups of constraints:

11.4.2 Passengers Constraints

$$\sum_{i \in I_w} h^w_i \leq d_w - \sum_{w' \in RE_w} g^w_{w'} + \sum_{w' \in RE_w} \alpha^w_{w'} g^w_{w'} \quad \forall w \in W \tag{11.4}$$

$$\sum_{i \in I_f} \sum_{w \in I_w} (1 - prob_i) h^w_i \leq \sum_{p \in P} alf_f q_p z^p_f \quad \forall f \in F \tag{11.5}$$
Constraints (11.4) state that the sum of all market passengers is lower than the market demand minus those passengers that the airline tries to disrupt to another market plus those passengers that come from another compatible market. These constraints account for attended, disrupted and recaptured passengers. Constraints (11.5) ensure that there are enough flight legs to satisfy passengers’ flows; misconnected passengers are removed from the flight leg.

### 11.4.3 Flight Leg and Airport Constraints

\[
\sum_{p \in P} z^p_f \leq 1 \quad \forall f \in F \tag{11.6}
\]

\[
\sum_{f \in F_{od}} \sum_{p \in P} z^p_f \geq m_{od} \quad \forall od \in ROD \tag{11.7}
\]

\[
\sum_{f \in CF_w} \sum_{p \in P} z^p_f \leq 1 \quad \forall w \in W \tag{11.8}
\]

\[
\sum_{p \in P} \sum_{f \in AS_{k,p}} z^p_f \leq qa_k \quad \forall k \in K \tag{11.9}
\]

\[
\sum_{p \in P} \sum_{f \in DS_{k,p}} z^p_f \leq qd_k \quad \forall k \in K \tag{11.10}
\]

Constraints (11.6) ensure that only one unique fleet type can be assigned to each flight leg. Constraints (11.7) state that for some origin-destination pairs a minimum service is required; the airline may be interested in that due to its strategy. Group of constraints (11.8) ensures that no more than one flight leg among those ones competing for passengers in the same market is scheduled; consequently, competition between flights from the same airline is avoided. Constraints (11.9)-(11.10) are airport capacity constraints that spare the departures and arrivals at airports at each period; this is mandated by the available slots in the airport to land or take off.

### 11.4.4 Fleet Flow, Capacity and Symmetry Constraints

\[
y^p_k - \sum_{f \in AS_{k,p}} z^p_f = y^p_k + \sum_{f \in DS_{k,p}} z^p_f \quad \forall k, p \in K, P \tag{11.11}
\]

\[
\sum_{f \in P} \beta_{f,t} z^p_f + \sum_{k \in K} y^p_k \leq N_p \quad \forall t, p \in CT, P \tag{11.12}
\]
\[
\sum_{f \in F} b_f z_f^p \geq f u_p \sum_{k_i \in K} y_k^p \quad \forall p \in P \tag{11.13}
\]
\[
y_k^p = y_{k_f} \quad \forall k, p \in K, P \tag{11.14}
\]

Block of constraints \((11.11)\) are the flow conservation equations for each airport and fleet type; they state that the inventory at some node is the inventory in the previous node minus departing planes plus arriving planes. Constraints \((11.12)\) are the fleet capacity constraints; we must count the necessary aircraft to perform the schedule and compare it to the available ones. Constraints \((11.13)\) ensure that the fleet average utilization is above a determined utilization rate provided by the airline; consequently, having an aircraft for just a few flights and a low utilization rate is avoided. Constraints \((11.14)\) state that the network must be symmetric in order to repeat the same schedule once the planning period has ended.

## 11.5 Case Study

As a proof of the model we have done some computational experiments in the network presented in \([10.3]\). The planning period we have considered is 24 hours. We require the planning to be periodic, that is, the fleet distribution must be equal at the beginning and the ending of the planning period. In this case time has been discretized into periods of 15 minutes. We have considered every potential flight leg between each spoke and the hub.

Using data provided by IBERIA, we compare the performance of the schedule provided by the integrated robust airline scheduling model to that of a non-robust approach.

### Computational results

The integrated robust airline scheduling model is implemented in GAMS 23.7 using ILOG CPLEX 12.2. Computational experiments are conducted on a workstation equipped with one Intel Intel Core2 Quad Q9950 2.83 GHz processor and 8 GB RAM.

The size of the model after CPLEX preprocessing is reported in Table \([11.1]\) for this study case.

As it was explained above, robustness is achieved through passengers that must perform flight connections. In order to demonstrate that a more robust schedule is obtained using the proposed approach, a comparison is made with a non-robust Integrated Airline Scheduling Model (IASM). The IASM is the same model explained above but removing
robustness aspects, that is, the objective function’s term penalizing misconnected passen-
gers, and the terms in constraints (11.5) accounting for misconnected passengers in flight
leg’s capacity. In order to measure the achieved robustness and compare both schedules
two different metrics are defined: the average connecting time per stopping passengers
($\bar{c}_{th}$) (11.15) and the average connecting time per itinerary fare ($\bar{c}_{tb}$) (11.16).

\[
\bar{c}_{th} = \frac{\sum_{w \in W} \sum_{i \in I_w} c_{ti} \cdot h_i^w}{\sum_{w \in W} \sum_{i \in I_w} h_i^w}
\] (11.15)
\[
\bar{c}_{tb} = \frac{\sum_{w \in W} \sum_{i \in I_w} c_{ti} \cdot b_i^w}{\sum_{w \in W} \sum_{i \in I_w} b_i^w}
\] (11.16)

where $c_{ti}$ is the available connecting time in itinerary $i$.

In Table 11.2, we show the differences between the robust and non-robust approach
regarding passengers. We report some numbers regarding the non-robust approach in the
second column and the robust approach in the third column. In the first row the total
number of attended passengers is shown; it is slightly lowered in the robust case because
some connecting passengers that had not enough connecting time are not attended (see
the third row where connecting passengers are shown). As there are fewer connecting
passengers there is room for attending more non-stop passengers (see the second row).
Misconnected passengers in the fourth row are clearly reduced achieving the goal we were
looking for. Recaptured passengers remain similar in both cases. Robustness metrics in
the sixth and seventh rows defined in (11.15) and (11.16) are increased; this means that
the connecting time is increased in the robust case achieving another objective of the
model; $\bar{c}_{tb}$ is slightly greater than $\bar{c}_{th}$ which means that the more profitable itineraries
are provided with greater connecting times. Finally, the average yield of the network
(revenue per passenger and kilometer) shown in the last row is increased in the robust
case; consequently, the more profitable passengers are protected in the robust case.

This reduction in the number of misconnected passengers will not be for free. The
price of robustness will remain in some other place. In order to analyze where the price of
Table 11.2 Passengers comparisons between the non-robust and robust approaches

<table>
<thead>
<tr>
<th></th>
<th>IASM</th>
<th>IRASMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attended passengers</td>
<td>6962.87</td>
<td>6799.42</td>
</tr>
<tr>
<td>Non-stop passengers</td>
<td>4853.63</td>
<td>4963.29</td>
</tr>
<tr>
<td>Connecting passengers</td>
<td>2109.24</td>
<td>1863.13</td>
</tr>
<tr>
<td>Misconnected passengers</td>
<td>270.41</td>
<td>152.12</td>
</tr>
<tr>
<td>Recaptured passengers</td>
<td>905.03</td>
<td>895.81</td>
</tr>
<tr>
<td>$\bar{e}_h$</td>
<td>85.407</td>
<td>99.031</td>
</tr>
<tr>
<td>$\bar{e}_b$</td>
<td>87.826</td>
<td>99.178</td>
</tr>
<tr>
<td>Average yield</td>
<td>0.0804</td>
<td>0.0820</td>
</tr>
</tbody>
</table>

Robustness remains, each of the costs in the objective function is shown in Table 11.3 for each of the approaches, IASM and IRASM. In the first row, the total revenue is shown. In the second row, costs of misconnected passengers are compared for the non-robust (IASM) and robust (IRASM) cases; for the robust case the cost is sensitively reduced. Robustness is achieved through the reduction in the expected number of misconnected passengers. Passengers’ costs with connection times ($\sum_{w \in W} \sum_{i \in I_w} \tau_i h^w_i$) are written in the third row. In the fourth row, passengers’ recapture costs are shown; these costs represent the lost revenue for attending passengers that were willing to fly in a determined market but finally they were recaptured in a different one. These costs are similar for both approaches. Operating costs are in the fifth row; they are lowered because there are fewer operated flights. In the last row, the objective function value appears for each case. In the robust case (IRASM), objective function’s value is lower than the non-robust one (IASM). The computational times in seconds for IASM and IRASM in the proposed network were 1124 and 1812, respectively.

Consequently, we can state that the robustness price remains in two different costs: the number of attended passengers that are slightly reduced and therefore the revenue is reduced; and the passengers’ cost with connecting time which is increased due to the fact that robustness is introduced by increasing connecting times.

In order to see how the robust model is scheduling flight connections we show an example in Figure 11.2. Here, we present a determined market, the market number 309. This market is defined by the origin SVQ and destination BIO and departure time 4:00 pm. Passengers must perform a flight connection in MAD. The IASM assigns passengers in the following way: 82.95 passengers to the itinerary composed of flights $k1 - k2$ and $k3 - k5$ (connecting time: 45 minutes) and 18.14 passengers to the itinerary composed of flights $k1 - k2$ and $k4 - k6$ (connecting time: 105 minutes). Consequently we have a total
Table 11.3 Costs comparisons between the non-robust and robust approaches

<table>
<thead>
<tr>
<th></th>
<th>IASM</th>
<th>IRASM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>661377.39</td>
<td>634362.54</td>
</tr>
<tr>
<td>Costs of Misconnected Passengers</td>
<td>38257.63</td>
<td>21855.96</td>
</tr>
<tr>
<td>Costs for Passengers with Connection Time</td>
<td>24019.29</td>
<td>24600.92</td>
</tr>
<tr>
<td>Recapture Costs</td>
<td>47494.96</td>
<td>48054.31</td>
</tr>
<tr>
<td>Operating Costs</td>
<td>495410</td>
<td>465355</td>
</tr>
<tr>
<td>Objective Function</td>
<td>94453.14</td>
<td>85424.31</td>
</tr>
<tr>
<td>Computational time</td>
<td>1124</td>
<td>1812</td>
</tr>
</tbody>
</table>

number of 20.19 misconnected passengers. However, the IRASM solution is as follows: 100.78 passengers to the itinerary composed of flights $k1 - k2$ and $k4 - k6$ and 0.31 to the itinerary composed of flights $k1 - k2$ and $k7 - k8$ (connecting time: 165 minutes). Therefore, we have a total number of 3.22 misconnected passengers.

We have also analyzed the fleet utilization. In Table 11.4 the used fleet resources are shown. As we can see, for the robust case (IRASM) an aircraft less is needed. The number in brackets in each element of the table shows the average block hour utilization for each fleet type in the proposed solution. As we can see there is on fleet type which is not used in this case study. IBERIA has these fleet types to attend all the European network. Consequently, we have enabled the model to choose the best types from those
Table 11.4 Fleet comparisons between the non-robust and robust approaches

<table>
<thead>
<tr>
<th></th>
<th>IASM</th>
<th>IRASM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-319</td>
<td>3 (8.16)</td>
<td>2 (8.74)</td>
</tr>
<tr>
<td>A-320</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A-321</td>
<td>11 (8.31)</td>
<td>11 (8.25)</td>
</tr>
</tbody>
</table>

Table 11.5 Fleet comparisons between the robust approaches with a minimum average block hour fleet utilization of 7 and 8

<table>
<thead>
<tr>
<th></th>
<th>IRASM(7)</th>
<th>IRASM(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-319</td>
<td>2 (7.08)</td>
<td>2 (8.74)</td>
</tr>
<tr>
<td>A-320</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A-321</td>
<td>13 (7.25)</td>
<td>11 (8.25)</td>
</tr>
</tbody>
</table>

to attend this simplified network. Regarding the results the fleet type A-320 should be removed from the fleet if just this simplified network was operated by IBERIA.

The number of used aircraft is lowered in the robust case. This is due to several issues: first, there are some passengers that are not ‘profitable’ anymore; enabling them a longer connection time would produce an increase in the cost perceived by them with the connecting time and more resources would be needed to do that; second, the lower bound in the fleet utilization means that canceling one single flight will probably produce the cancellation of more flights. Consequently, getting lowered the fleet utilization will enable the model to provide solutions with additional aircraft. Table 11.5 shows that by getting lowered the minimum average fleet utilization to seven block hours additional aircraft are used in the solution. The second column shows the robust solution for a minimum average fleet utilization of seven block hours (IRASM(7)) and the third one for 8 block hours (IRASM(8)).

In order to analyze the new solution quality regarding passengers Table 11.6 shows the same information as Table 11.2 but for the robust cases with a minimum average fleet utilization of seven block hours (IRASM(7)) and eight block hours (IRASM(8)). As it was expected, more passengers are attended in the solution where a minimum average fleet utilization of seven block hours is imposed. Misconnected passengers remain similar in both cases. The same applies to the connecting time metrics ($\overline{ct}_b$ and $\overline{ct}_b$). Recaptured passengers number is increased because in the new solution two more aircraft of the biggest type are scheduled and passengers are aggregated in them.

179
Table 11.6 Passengers comparisons between the robust approaches with a minimum average block hour fleet utilization of 7 and 8

<table>
<thead>
<tr>
<th></th>
<th>IRASM(7)</th>
<th>IRASM(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attended passengers</td>
<td>6902.99</td>
<td>6799.42</td>
</tr>
<tr>
<td>Non-stop passengers</td>
<td>5036.23</td>
<td>4963.29</td>
</tr>
<tr>
<td>Connecting passengers</td>
<td>1866.76</td>
<td>1863.13</td>
</tr>
<tr>
<td>Misconnected Passengers</td>
<td>153.01</td>
<td>152.12</td>
</tr>
<tr>
<td>Recaptured Passengers</td>
<td>979.99</td>
<td>895.81</td>
</tr>
<tr>
<td>$\bar{c_{h}}$</td>
<td>98.26</td>
<td>99.031</td>
</tr>
<tr>
<td>$\bar{c_{b}}$</td>
<td>97.92</td>
<td>99.178</td>
</tr>
<tr>
<td>Average yield</td>
<td>0.0821</td>
<td>0.0820</td>
</tr>
</tbody>
</table>

11.6 Summary

We have proposed a new robust approach to solve the airline scheduling problem, where schedule design and fleet assignment problems are jointly solved. In addition, passengers’ flows are obtained through different itineraries in the network accounting for the connection time.

The model has been tested in a IBERIA’s simplified network. Computational results show how robustness may be achieved. However, this robustness has a price. The robust approach has been compared with a non-robust approach showing the price of the achieved robustness.
Chapter 12

INTEGRATED AIRLINE SCHEDULING: CONSIDERING COMPETITION EFFECTS AND THE ENTRY OF THE HIGH SPEED RAIL

Airlines have been using operations research techniques since the 1950s but we can point to major applications since the late 1970s, when the United States airline industry was deregulated. Competition between airlines affects the number of captured passengers and, therefore, revenues. Moreover, the airline industry is now facing a new hassle: High-speed rail. They are increasingly competing for passengers. Therefore, we study multimodal competition including airline and high-speed rail, and develop a new approach that estimates the demand associated with a given schedule using a nested logit model, and generates airline schedules using an integrated optimization model that captures frequency planning, approximate timetable development, fleet assignment and passenger demand choice. Our experimental results show that we are able to replicate IBERIA’s (the major Spanish airline) current decisions with a reasonable level of accuracy, thus validating our modeling approach. We evaluate multiple scenarios involving entry of High-speed rail in some markets, and we account for the possibility of demand stimulation as a result of the new services.
12.1 Introduction

Airline success depends on creating and delivering products and services valued by customers. Historically airlines have used multiple models to analyze the effects of schedule, price and competition. However, they have not been integrated with schedule development models. Given the high level of competition in the airline industry, effective decision tools are crucial to the profitability of an airline.

In scheduled air transportation, airline profitability is critically influenced by the airline’s ability to:

1. estimate passenger demands;

2. construct profitable flight schedules (referred to as the airline schedule planning process);

3. and determine the fare levels for a set of products in an origin-destination market (referred to as the pricing process) and determine how many seats to make available at each fare level (referred to as revenue management).

12.1.1 Demand Estimation

In practice, an airline can usually estimate its competitors’ services for the next season within a reasonable level of accuracy. Planners from the marketing department estimate the market share and passenger demand for the next season based on the projected future market demand, as well as its competitors’ current and past operational data, including market share, passenger demand, timetable and other services.

Demand models are used to develop forecasts of passenger demand for each origin-destination (OD) pair (or market) as a function of attributes such as average fares, frequencies, market demographics and economic conditions (Garrow [68]). In the past studies on airline scheduling models, the market demand for a specific origin-destination (OD) pair is usually assumed to be fixed. (Dumas et al. [60] present an improved fleet assignment model that takes stochastic demand prediction as inputs, and aims at computing expected numbers of passengers on each itinerary. However, they do not consider competition). This assumption is reasonable when the offered service is assumed to be regular, i.e. no entry by a new operator is expected. The entry by new operators into markets will lead to a variation in average price and frequency while also stimulating passenger demand and thus varying the total volume of demand.
Given these total demand estimates, passenger choice models are used to estimate for each competitor and each market, the proportion of demand (or share) it captures in that market, considering market-specific characteristics, including passengers’ mode preferences and airline preferences, fares, flight frequencies and other market characteristics.

There is widespread acceptance in the airline industry of an S-curve relationship between airline market share and frequency share. The S-curve describes how an airline’s market share grows non-linearly with its frequency share in that market. Early theoretical development and empirical evidence that higher-frequency shares are associated with disproportionately higher market shares was provided in the 1970s before deregulation (Simpson [113]). After deregulation, there exist a number of references to the S-curve (Wei and Hansen [125], Belobaba [15], Vaze [121], Vaze and Barnhart [122]).

The most commonly used mathematical expression for the S-curve relationship is given by:

\[
P^w_a = \frac{[f^w_a]^\alpha}{\sum_{a' \in A} [f^{w'}_{a'}]^\alpha},
\]

where \(A\) is the set of airlines, \(P^w_a\) is the probability that a passenger in market \(w\) selects airline \(a\) among all the airlines, \(f^w_a\) is the frequency value of airline \(a\) in market \(w\) and \(\alpha\) is the frequency parameter.

### 12.1.2 Schedule Planning

The schedule planning process typically starts from an existing schedule with a well-developed route structure and fleet composition (see 10.2). In constructing each new schedule, changes are introduced to the existing schedule to reflect changes in demands and the environment. Due to the enormous size and complexity of the problem, schedule planning is a multi-step process, usually separated into four, sequentially solved subproblems: schedule design, fleet assignment, maintenance routing, and crew scheduling (Barnhart and Cohn [9], Barnhart et al. [12]).

For tractability purposes, schedule design models typically consider demand for an airline’s flights to be deterministic and invariant to schedule changes and competition. These assumptions, however, have been shown to lead to overestimates of the number of passengers served, the revenue captured, and schedule profitability (Yan et al. [129], Belobaba [15]). An effective schedule planning process for an airline, then, depends critically on both the accurate estimation of the overall demand for travel on itineraries.
between markets; and the accurate understanding of how passengers will choose between the airline’s and its competitors’ travel options.

12.1.3 Schedule Planning and Demand Estimation

Improved plans can be generated solving integrated models of some the planning phases such as schedule design and fleet assignment. To capture more accurately the supply and demand interaction more accurately, we integrate the schedule design and fleet assignment modeling steps in our approach. In the basic fleet assignment model formulations important network considerations are insufficiently treated and the resulting solutions are often suboptimal. Barnhart et al. [7] propose a new formulation and solution approach that captures network effects and generates superior solutions. They propose an itinerary-based model formulation. Jacobs et al. [76] incorporate origin and destination network effects into the fleet assignment process combining a modified version of a leg-based fleet assignment model with the network flow aspects of probabilistic yield management. Using a passenger flow model devised in Dumas and Soumis [59], Dumas et al. [60] improve the fleet assignment model by taking stochastic demand prediction as inputs, and aim at computing expected numbers of passengers on each itinerary. Barnhart et al. [12] employ composite decision variables representing the simultaneous assignment of fleet types to subnetworks of one or more flight legs. The formulation is motivated by the need to better model the revenue side of the objective function. Vaze and Barnhart [123] solve the aggregated timetable development and fleet assignment problem to minimize the system-wide delays assuming a single monopolistic carrier that satisfies all the passenger demand in the USA. They develop a bound on the minimum possible level of delays that can be achieved. Cadarso and Marín [37] propose an integrated approach that optimizes the schedule design, the fleet assignment and the passenger demand to produce more economical solutions and create fewer incompatibilities between the decisions. They provide robust itineraries in order to ameliorate disconnected passengers and report computational tests on realistic problem instances of the Spanish major airline IBERIA.

However, none of the aforementioned references account for passenger demand behavior and competition among airlines to capture the demand. A carrier cannot neglect the influence of its schedule on its market share. The passenger demand for an airline could be adversely impacted if the schedule is considered to be less attractive by the passengers. On the other hand, a good schedule that takes into consideration passenger reactions to its services attracts more passengers and improves the carrier’s market share in actual operations. Therefore, to set a good schedule, not only does the fleet and related supply
have to be considered, but also variable market shares in a competitive market have to 
be taken into account. Therefore, we include competitive effects in our model because 
schedule design models that do not include competition result in overestimates of revenues 
due to overstatement of passengers captured (Yan et al. [129], Belobaba [15]).

An essential component of our approach is to capture variation in demand with changes 
in schedules. Hansen [71] develops a model of airline hub competition and considers 
airline competitors belonging to two types: hub carriers and direct carriers. Hong and 
Harker [75] present a model which is aimed at developing an efficient air traffic system 
for given demand and airport capacity levels by the proper pricing of landing slots. A 
computable Nash equilibrium model is used in the context of a two-stage, game-theoretic 
representation of a market mechanism for slot allocation. Yan et al. [129] develop a short-
term flight scheduling model with variable market shares in order to help a Taiwanese 
airline to solve for better fleet routes and flight schedules in competitive markets. The 
modeling approach is limited in that they do not account for itineraries, that is, the model 
is arc based. Wei and Hansen [126] address the question of how airlines make decisions on 
aircraft size and service frequency in a competitive environment. They apply three game-
theoretic models to analyze how airlines’ choices in a competitive environment may vary 
with flight distance. Pita et al. [104] present a model whose objective is to maximize the 
expected profits of an airline that faces a given origin/destination-based travel demand 
and operates in congested, slot-constrained airports. Both airline competition and airline 
cooperation are dealt with in the model, though in a simplified manner (i.e., in terms of 
demand modeling frequency is the only attribute used).

These references capture variation in demand with changes in schedules but they do 
not make differentiation among Legacy Airlines (LA) and Low Cost Airlines (LCA) (LAs 
and LCAs provide different service levels, and hence, the resulting demand patterns are 
different). Moreover, there is only one transportation mode available. Our approach 
deals with different modes of transportation and differentiates among different types of 
operators. In a particular market, passengers have to first choose the mode of transport. 
In short-haul passenger markets, High Speed Rail (HSR) and airlines are increasingly 
competing for passengers, especially in many parts of Europe and Asia. HSR plays a 
significant role in the short- to medium-haul markets. It can shorten travel times between 
cities and their comparative advantages against other competing modes are based on 
quality of service. The expansion of HSR around the world and its dominance, particularly 
in direct services, calls for the inclusion of inter-modal competition.

To more accurately estimate the revenues associated with an airline flight schedule, we
develop an airline schedule design model that captures the impacts on passenger demand of schedule decisions, as suggested by the S-curve relationship which is a popular notion in the airline industry (Belobaba [15], Simpson [113]). Vaze and Barnhart [122] state that airline frequency competition is partially responsible for the growing demand for airport resources. They propose a game-theoretic model for airline frequency competition modeled according to the S-curve relationship between the market share and frequency share. The model is solved to obtain a Nash equilibrium using a successive optimizations approach. The model predictions are validated against actual frequency data, with the results indicating a close fit to reality.

A critical capability is to estimate the demand associated with a schedule. We develop a logit model to estimate demand for different schedules and we evaluate our approach with extensive computational experiments using data representing air and rail transportation options in a particular country in Europe. We consider multi-modal competition: high speed rail and other airlines, differentiating between low-cost and legacy airlines. Previous econometric studies have investigated this multi-modal competition using logit models to estimate the demand associated to with a schedule: Behrens and Pels [14] study inter-modal competition in the London-Paris passenger market. Using revealed preference data, they estimate nested and mixed multinomial logit models to examine passenger behavior in the London-Paris market. Román et al. [105] analyze the potential competition of the HSR with the air transport between Madrid and Barcelona in Spain. The analysis estimates disaggregated mode choice models using information provided by mixed revealed and stated preferences database. However, there is limited research about integrating this type of multi-modal logit model with a schedule design model. Zito et al. [130] model how airlines make operative decisions on fares and frequencies of service in a competitive environment. They investigate short haul market for intercity linkages. In this segment the air mode is in competition with other ground modes. Although the authors include mode competition, they do not include competition between LA and LCA. Moreover, the schedule development is not included in detail and no real study case is presented.

Our work differs from others in that it considers multi-modal competition including airline (legacy and low-cost) and high-speed rail, and develops a new approach that: 1) estimates the demand associated with a given schedule using a logit model; and 2) generates airline schedules and fleet assignments using an integrated schedule design and fleet assignment optimization model that accounts for passenger demand competition, and captures the impacts of schedule decisions on passenger demand, as suggested by the S-curve relationship. Our model, reflecting that passenger demand for an airline schedule
depends not only on the airline’s schedule but also on the schedule of its competitors, captures linkages between schedule competition and passenger demand and is therefore able to drive profit maximization with improved estimates of revenues. However, pricing and revenue management decisions are out of the scope of our model, that is, the model uses average ticket fares as inputs.

The motivation for considering multi-modal competition stems from the fact that high-speed rail (HSR) and airlines are increasingly competing for passengers in many parts of Europe and Asia, especially in short- to medium-haul markets. HSR competes by often providing similar or even greater service frequency and better connectivity to the city centers. Moreover, HSR is often perceived as the safer and more comfortable mode. In addition to modeling competition between air and rail modes, we model competition between legacy and low-cost airlines. These are perceived as different choices because level of service is different. Legacy airlines include the following services: first class and/or business class, a frequent-flyer program, airport lounges, alliance partners that agree to provide these services to the passengers as well, etc.

The remainder of this chapter is organized as follows. In Sections 12.2 and 12.3, respectively, we present our demand modeling approach and our schedule design under multi-modal competition formulation. In Section 12.4, computational experiments for a real-world problem using data provided by a legacy airline in Europe are described, and we detail our results.

12.2 Demand Modeling

Apart from fare, service frequency is the most important attribute on which the airlines compete. An airline can attract more passengers in a market (defined by the departure airport, arrival airport and a departure time period) by increasing the frequency on a route. For a given unconstrained total demand on a route, the market share of each airline depends, among other factors, on its own frequency and on the frequency of its competitors. Many past studies have modeled market share as a function of frequency share using an S-shaped or sigmoidal relationship (Vaze and Barnhart [122]).

However, modeling the market share as simply a function of the frequency share is not enough to model passenger demand behavior in many markets. This is especially true in markets where the competitor fares are different from each other and the competing airlines are different from the perspectives of the passengers in other ways (Vaze and Barnhart [122]). There are other attributes, such as fares and travel times that can
substantially affect passengers’ airline choice (Behrens and Pels [14], Román et al. [105]). For instance, consider an Origin-Destination (OD) pair operated by two different airlines: the first one operates a direct flight and the second one a one-stop flight. Passengers will likely prefer the direct flight over the one-stop flight, all else being equal. Consequently, we extend the model in 12.1 in order to include fare and travel time as attributes as follows:

$$P_w^a = \frac{asc_a \left[ f_w^a \right]^\alpha \left[ p_w^a \right]^\beta \left[ m_w^a \right]^\gamma}{\sum_{a' \in A} asc_{a'} \left[ f_{w}^{a'} \right]^\alpha \left[ p_{w}^{a'} \right]^\beta \left[ m_{w}^{a'} \right]^\gamma},$$  

(12.2)

where $p_w^a$ is the price of airline $a$ in market $w$, $m_w^a$ is a measure of flight time of airline $a$ in market $w$ (for both direct flights and connecting flights; for connecting flights we use an average connecting time), $\beta$ is the price parameter and $\gamma$ is the flight time parameter. In addition, there could be other airline-specific factors that impact the passenger share. For example, some passengers might have a preference for legacy carriers over low cost carriers, or some passengers might prefer one airline over the other due to frequent flyer program memberships, etc. In order to capture these factors, we include alternative specific constants for each airline $asc_a$.

In general, all passengers prefer lower fares and higher frequency. However, some passengers might value lower fares more than other passengers, while others might give more importance to higher frequency and the associated greater flexibility in scheduling their travel. To incorporate these effects, we propose an extension of 12.2. Let $Z$ be the set of passenger types and $\varsigma \in Z$ represent passenger types business or leisure. Let $\gamma_w^\varsigma$ be the fraction of market $w$ passengers belonging to type $\varsigma$ such that $\sum_{\varsigma \in Z} \gamma_w^\varsigma = 1$. Consequently, the probability for a passenger of type $\varsigma$ in market $w$ of selecting airline $a \in A$ among all the airlines will be as follows:

$$P_w^{a,\varsigma} = \frac{asc_a^\varsigma \left[ f_w^a \right]^\alpha^\varsigma \left[ p_w^a \right]^\beta^\varsigma \left[ m_w^a \right]^\gamma^\varsigma}{\sum_{a' \in A} asc_{a'}^\varsigma \left[ f_{w}^{a'} \right]^\alpha^\varsigma \left[ p_{w}^{a'} \right]^\beta^\varsigma \left[ m_{w}^{a'} \right]^\gamma^\varsigma},$$  

(12.3)

where $\alpha^\varsigma$ is the frequency parameter, $\beta^\varsigma$ is the price parameter, $\gamma^\varsigma$ is the flight time parameter and $asc_a^\varsigma$ is the alternative specific constant for each airline for each passenger type $\varsigma$.

During the last few decades in Europe, High Speed Rail (HSR) has become an important competitor, presenting railway transport in a new form and notably improving the
quality of service offered. Behrens and Pels [14] and Román et al. [105] present two
different case studies in Europe. An investment in HSR often generates a redistribution
of passengers between air and rail alternatives, varying the existing modal distribution.
The impact of this investment on the demand for airline travel is quite uncertain and of
great interest. For some, HSR is considered the best transport mode for short distance
trips, providing shorter travel times between cities, higher quality of service and reduced
access times to city centers.

We, therefore, introduce a modal choice model. Passengers must decide between rail
and air modes for the OD pairs where both are competing. We assume that each operator
offers a unique route for each OD pair. The choice by a representative passenger of a
specific alternative in each OD pair can be modeled using the nested multinomial logit
model. We have two levels in the model: mode (air vs. rail) and operator. We model
passengers’ modal choice in

\[
P_{\text{air}} = \frac{\left( \sum_{a \in A} \frac{\text{asc}_a \left( f_w^a \right)^{\alpha_a} \left( p_w^a \right)^{\beta_w} \left( m_w^a \right)^{\gamma_w}}{\text{asc}_a \left( f_r^a \right)^{\alpha_r} \left( p_r^a \right)^{\beta_r} \left( m_r^a \right)^{\gamma_r}} \right)^{\theta_f} \left( t_{w,\text{air}} \right)^{\gamma_f}}{\left( \sum_{a \in A} \frac{\text{asc}_a \left( f_w^a \right)^{\alpha_a} \left( p_w^a \right)^{\beta_w} \left( m_w^a \right)^{\gamma_w}}{\text{asc}_a \left( f_r^a \right)^{\alpha_r} \left( p_r^a \right)^{\beta_r} \left( m_r^a \right)^{\gamma_r}} \right)^{\theta_r} \left( t_{w,\text{air}} \right)^{\gamma_r}} + \text{asc}_a \left( f_w^a \right)^{\alpha_a} \left( p_w^a \right)^{\beta_a} \left( m_w^a \right)^{\gamma_a} \right],
\]

where \(\theta_f\) is a measure of the degree of correlation and substitution among alternatives
in the nest. The lower the correlation between unobserved effects in the utilities of different
air operators, the higher is the \(\theta_f\) value. \(\text{asc}_a\) is the rail alternative specific constant. \(t_{w,\text{air}}\) and \(t_{w,\text{rail}}\)
are travel times for the air and rail modes in market \(w\), respectively. \(\gamma_f\) and \(\gamma_r\) are
travel time parameters. \(f_w\) is the rail service frequency and \(\alpha_r\) is the frequency parameter
in the rail mode. \(p_w\) is the rail mode’s price and \(\beta_r\) is the price parameter for the railway.
The probability for a passenger of selecting the rail mode is: \(P_{\text{rail}} = 1 - \sum_{a \in A} P_{a|\text{air}} P_{\text{air}}\)
, where \(P_{a|\text{air}}\) is the expression in [12,3]. Consequently, the probability for a passenger in
this multimodal context of selecting airline \(a\) is: \(P_{a|\text{air}} = P_{a|\text{air}} P_{\text{air}}\).

In the rail mode, there is no differentiation in passenger types. This is a limitation
of our analysis due to a lack of data regarding passenger types in HSR in our case study
data. Moreover, we must note that in the markets where the HSR operates, there are no
connecting flights in our case study. Consequently, the attribute \(m_w\) is redundant and
hence, it is not included.
12.2.1 Nested Logit Model Parameter Estimation

We have estimated the nested logit model parameters using real data from year 2010 provided by IBERIA. Since we are estimating this as a discrete choice model, we rewrite the expressions in 12.3 and 12.4 (see 12.5) so that we can show the standard discrete choice formulation where the choice probability is the ratio of the exponentials of the systematic utilities. Utility is a function of systematic and random components with the systematic component being a linear combination of logarithms of frequency, fare, travel time, and alternative specific constants (the random part of the utility function is assumed to be independent and identically Gumbel distributed). The parameters to be estimated are: $\alpha^\varsigma$, $\beta^\varsigma$, $\gamma^\varsigma$, $asc^\varsigma_a$, $\theta_f$, $asc_f$, $\gamma_f$, $\gamma_r$, $\alpha_r$ and $\beta_r$. The data are composed of 104 different origin-destination pairs and 7 different periods of time. Consequently, there are 728 different sets of data composed of: market share, frequencies, itinerary characteristics, travel time and average price. Each element on the set corresponds to a time period of one week. Each set is composed of the data of the airline under study (IBERIA) and its competing airlines. There are two main types of competing airlines: the low cost and the legacy airlines. There are three low cost airlines and three legacy airlines in this case study. The rail competition is provided by a unique rail operator and it is not present in all the markets: the HSR is present in the following OD pairs: Madrid-Barcelona, Barcelona-Madrid, Madrid-Sevilla, Sevilla-Madrid, Madrid-Málaga and Málaga-Madrid.

There are two types of passengers, indexed by $\varsigma \in \mathbb{Z}$ with $\varsigma$ equal to 1 representing business passengers and $\varsigma$ equal to 2 representing leisure passengers. As for the alternative specific constant, $asc^\varsigma_a$, we have used it to model the passenger perception of legacy and low cost airlines. Thus, it takes value 1 if the airline is a legacy airline and it takes the value in Table 12.1 if the airline is a low cost airline.

We have used the maximum likelihood estimation method in order to estimate the nested logit model parameters. The Newton-Raphson method is used to maximize the likelihood function with respect to the parameter vector. All of the parameter estimates are significant at the 0.95 confidence level using a classic Student’s t-test. Table 12.1 shows the test results: the parameter to be estimated in the first column, the estimation in the second column, the standard error in the third column and the p-values in the last column.

We use the log-likelihood ratio-test statistic, following a Chi-squared distribution with degrees of freedom equal to the number of extra estimated parameters to test overall model significance. This test statistic equals 1011.32. Using the critical value of 19.675 (eleven extra parameter), we reject the null hypothesis that the naive model is better.
Table 12.1 Estimation results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimation</th>
<th>Std error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha^1$</td>
<td>1.5137</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td>$\beta^1$</td>
<td>-0.4186</td>
<td>0.035</td>
<td>0.008</td>
</tr>
<tr>
<td>$\gamma^1$</td>
<td>-1.5050</td>
<td>0.042</td>
<td>0.017</td>
</tr>
<tr>
<td>$\text{asc}_a^1$</td>
<td>0.8265</td>
<td>0.052</td>
<td>0.023</td>
</tr>
<tr>
<td>$\alpha^2$</td>
<td>1.18201</td>
<td>0.021</td>
<td>0.010</td>
</tr>
<tr>
<td>$\beta^2$</td>
<td>-0.9347</td>
<td>0.063</td>
<td>0.012</td>
</tr>
<tr>
<td>$\gamma^2$</td>
<td>-1.3251</td>
<td>0.112</td>
<td>0.017</td>
</tr>
<tr>
<td>$\text{asc}_a^2$</td>
<td>1.0853</td>
<td>0.047</td>
<td>0.009</td>
</tr>
<tr>
<td>$\theta_f$</td>
<td>0.6583</td>
<td>0.073</td>
<td>0.019</td>
</tr>
<tr>
<td>$\gamma_f$</td>
<td>-0.8568</td>
<td>0.129</td>
<td>0.031</td>
</tr>
<tr>
<td>$\alpha_r$</td>
<td>0.9195</td>
<td>0.109</td>
<td>0.025</td>
</tr>
<tr>
<td>$\beta_r$</td>
<td>-0.4451</td>
<td>0.093</td>
<td>0.018</td>
</tr>
<tr>
<td>$\gamma_r$</td>
<td>-0.8844</td>
<td>0.184</td>
<td>0.037</td>
</tr>
<tr>
<td>$\text{asc}_r$</td>
<td>1.4343</td>
<td>0.152</td>
<td>0.034</td>
</tr>
</tbody>
</table>

than the nested logit model.

The estimated inclusive value parameter of the air nest ($\theta_f$) is far from one. The interpretation is that, for the presented study case, air alternatives cannot be considered as substitutes in unobserved effects (the log-likelihood ratio-test statistic equals 14.89; using the critical value of 3.84 (one extra parameter), we reject the null hypothesis that the model with $\theta_f = 1$ is better than the nested logit model).

From the alternative specific constant for the railway, we can conclude that, all else being equal, passengers prefer railway rather than the air mode when competing in the same origin-destination pair ($\text{asc}_r > 1$); this is mainly due to unobserved effects such as the fact that the railway operates between city centers, railway security checks are less onerous for the passengers, and rail passenger travel is more comfortable and convenient (allowing the use of electronic devices throughout the journey, etc.).

From the estimation results we conclude that passengers’ value of time is similar for different transport modes ($\gamma_f, \gamma_r$). However, passengers value railway frequency less than airline frequency ($\alpha_r < 1$ and $\alpha^1$ and $\alpha^2 > 1$). This might be due to the fact that railways are perceived by passengers as more reliable systems or that rail travel time can be more productively utilized by passengers. HSR passengers’ sensitivity to prices ($\beta_r$) is similar to the price sensitivity of business-category air travelers.

In our two stage model, we assume that the passenger first picks the mode and then decides between different air alternatives if the air mode is chosen. The results show that business passengers prefer legacy airlines over low cost airlines ($\text{asc}_a^1 < 1$), while leisure
passengers prefer low cost airlines \((asc_a^2 > 1)\). This suggests that in addition to the attributes we captured in our model (that is, fare, frequency and travel time) there are other attributes, such as differences in airline image, loyalty programs etc., that make low cost airlines relatively more attractive to leisure passengers and legacy airlines relatively more attractive to business passengers.

Passengers belonging to the business category are more sensitive to travel time than passengers belonging to the leisure category \((\gamma^1 < \gamma^2)\), while business passengers are less sensitive to higher fares than leisure passengers \((\beta^1 > \beta^2)\). In addition, business passengers place more value than leisure passengers on higher frequency values \((\alpha^1 > \alpha^2)\), presumably because higher frequency reduces effective travel time, which includes schedule displacement Belobaba [15]. Schedule displacement corresponds to time elapsed between when a passenger wants to travel and when service is offered. Thus, effective travel time and schedule displacement both change inversely with frequency.

### 12.2.2 Captured Demand by an Airline

The unconstrained demand of passengers \(d_w\) in each market \(w\) is assumed to be fixed. The captured demand by airline \(a\) depends on competition effects, as explained in Section 12.2. Consequently, unconstrained demand gets split between different modes and operators. The captured demand \((\Delta_w^a)\) by airline \(a\) in market \(w\) is \(\Delta_w^a = \sum_{c \in Z} P_{a|w}^c P_{w|a}^c \gamma_w d_w\).

As we are studying a tactical problem for which the timetable is approximate, we model captured demand as a function of the total frequency offered in the OD pair. We therefore, account for OD average attributes in the logit model formulation instead of market specific attributes. Consequently, \(od\) represents the requested OD pair by market \(w\). Hence, the captured demand depends on the total frequency values and average prices in the origin-destination pair during the entire planning period. The extended formulation, which has been modified as explained, is in 12.5.

\[
\Delta_w^a = \sum_{c \in Z} \left[ e^{log(asc_{w,c}^a) + \alpha^c log(fod_{w,c}^a) + \beta^c log(pod_{w,c}^a) + \gamma^c log(mod_{w,c}^a)} \cdot \left( \frac{\theta f log \left( \sum_{a' \in A} e^{log(asc_{w,c}^{a'}) + \alpha^c log(fod_{w,c}^{a'}) + \beta^c log(pod_{w,c}^{a'}) + \gamma^c log(mod_{w,c}^{a'})} \right) + \gamma f log(ttod_f)}{\epsilon \left( \sum_{a' \in A} e^{log(asc_{w,c}^{a'}) + \alpha^c log(fod_{w,c}^{a'}) + \beta^c log(pod_{w,c}^{a'}) + \gamma^c log(mod_{w,c}^{a'})} \right) + \gamma f log(ttod_f)} \right) \right] d_w
\]

192
12.3 Optimization Model Formulation

Vaze and Barnhart [123] present a three-stage modeling framework for the airline planning process. The first stage deals with network design where decisions about the number and location of hubs, candidates for non-stop routes and allowable airports for passengers’ connections are made. The second stage involves the daily frequency planning and fleet assignment problem, and the third stage addresses the timetable development and fleet balancing problem. The model presented in this paper combines the decisions in the second and third stages of this framework. The integrated model addresses frequency planning, fleet assignment and development of an approximate timetable (i.e., number of departures per airport-time period (see Air Network)). As we are proposing a tactical competition model, we assume that fares are fixed at an average value and will not be part of the decision variables. We will formulate the model as an optimization problem from an airline’s point of view. Let us consider airline \( a \).

**Air Network.** The air network is formed by the airports and all the feasible airway alternatives linking them. The airports are characterized by the operations that can be performed at those airports. We assume that landing and departure slot availability is known for each airport and airline and airways are the links (flight opportunities) between the airports.

In order to ensure tractability of the problem, we propose an aggregated network, as presented in Harsha [73]. It allows different levels of time-discretization at each airport, depending on the level of airport congestion. More congested airports are modeled with a finer-level of discretization to ensure that the number of flight operations does not exceed capacity at any (short) period of time. Uncongested airports are modeled with a lower level of fidelity, as capacity constraints are not typically binding. Such an aggregation scheme reduces problem size smaller compared to that using a single discretization level.

When an airway is selected it is referred to as a flight leg, defined by a departure airport-time period and arrival airport-time period. An airport-time period is a combination of an airport and a specific time period at that airport. There are different fleet types, with each fleet type primarily characterized by its seating capacity. Our optimization formulation allows for practical constraints that ensure that certain aspects of an existing flight schedule are included in the new schedule. For example, an airline sometimes receives government grants to maintain a minimum level of service in some markets; an airline sometimes schedules a pre-determined minimum number of flights in some OD pairs in order to maintain its competitive position in those markets; and an airline can lose some of its departure and landing slots at an airport if some OD pairs are
not operated.

Finally, we assume that the flight schedule will be periodic, that is, the schedule will repeat after the planning period ends. To satisfy this, the locations of aircraft types at the beginning and the end of the planning period must be the same.

**Passenger Demand.** Unconstrained demand is estimated for the combination of origin airport, destination airport and desired departure time period. This combination is defined as a market. Consequently, for each market the departure airport-time period is known. The unconstrained total demand of passengers in each market is assumed to be fixed and known. For our study an estimate of such demand was provided by the airline. An important modeling issue is how to capture the passenger demand satisfaction requirements for every OD pair and every time period of the day Vaze and Barnhart [123]. To model this, we divide the day into time periods in the same way as for airports and ensure that passengers are carried during their desired time period.

In each market, passengers can choose any of the corresponding itineraries. Each itinerary is defined by a set of flight legs that connect the origin and destination airports. It can be composed of one flight leg or it can consist of more than one leg including intermediate stops at different airports. Thus the proposed model is itinerary-based. The demand captured by a certain airline will depend on competition effects which are incorporated in our model through the nested logit model introduced in section [122].

As mentioned above, we are integrating several sub-problems namely, frequency planning, approximate timetable development and fleet assignment, all while considering passenger demand variation with schedule. Then, the aim of our Integrated Airline Scheduling under COmpetition Model (IASCOM) is to obtain, for all flights for one single airline \(a\), the frequency by fleet type and by time period in the day, given airport capacities, fleet sizes, average fares, unconstrained demands and competitors’ schedules.

The notation in the IASCOM is defined as follows:

Sets:

- \(A\): set of operators indexed by \(a\).
- \(G\): set of airports indexed by \(g\).
- \(OD\): set of origin-destination pairs indexed by \(od\).
- \(W\): set of markets indexed by \(w\); each element in the set is defined by an origin-destination pair and a departure time period.
• $I$: set of itineraries indexed by $i$; each element is defined by an origin-destination pair, and the flight legs connecting it.

• $F$: set of flights indexed by $f$; each element is defined by an origin-destination pair, a departure time period and an arrival time period.

• $\Pi$: set of fleet types indexed by $\pi$.

• $K$: set of nodes indexed by $k$; a node is an airport at a point in time.

• $Z$: set of passenger types indexed by $\varsigma$. There are two types of passengers, with $\varsigma$ equal to 1 representing business passengers and $\varsigma$ equal to 2 representing leisure passengers.

• $K_g$: subset of nodes belonging to airport $g$.

• $PK_k$: subset of nodes including $k$ and all those that precede it in time at the same airport.

• $I_w$: subset of itineraries serving market $w$.

• $I_f$: subset of itineraries using flight $f$.

• $AF_k$: subset of flights arriving in node $k$.

• $DF_k$: subset of flights departing from node $k$.

• $F1_{od}$: subset of flights which serve $od$ pair as the first flight leg in the itinerary.

• $F2_{od}$: subset of flights which serve $od$ pair as the second flight leg in the itinerary.

• $RF_f$: subset of flights that are subject to some regulation.

Parameters:

• $p_w$: average ticket price for market $w$.

• $c_f^\pi$: operating cost for flight $f$ with fleet type $\pi$.

• $d_w$: unconstrained demand in market $w$.

• $lfm_f$: maximum allowable load factor for flight $f$.

• $q_\pi$: seating capacity of fleet type $\pi$. 

195
• $qa_k^a$: available arrival slots in node $k$ for airline $a$.
• $qd_k^a$: available departure slots in node $k$ for airline $a$.
• $\tau_f$: block time for flight $f$.
• $h_\pi$: average block time for fleet type $\pi$.
• $n_\pi$: fleet size for fleet type $\pi$.
• $m_f$: minimum frequency to be operated in flights subject to some regulation.
• $u_\pi^g$: maximum number of planes of fleet type $\pi$ grounded at airport $g$.
• $P_{\text{air}}^{w,\varsigma}$: probability for a passenger of type $\varsigma$ from market $w$ of selecting nest $\text{air}$ among all the nests.
• $P_{\text{air}}^{w,\varsigma}$: probability for a passenger of type $\varsigma$ from market $w$ of selecting alternative $a$ among all the alternatives in the nest.
• $\Delta_w^a$: number of passengers captured by airline $a$ in market $w$.
• $\gamma_\varsigma^w$: proportion of passengers of type $\varsigma$ in market $w$.

Variables:
• $z_\pi^f$: frequency of flight $f$ with fleet type $\pi$.
• $y_\pi^g$: number of planes of fleet type $\pi$ on the ground at the beginning of the planning period in airport $g$.
• $h_i$: passengers flown on itinerary $i$.
• $f_{a,od}^i$: frequency offered by airline $a$ in $od$ pair.

The mathematical model for an operator (an airline) $a$ is as follows:

$$\max z_a = \sum_{w \in W} \sum_{i \in I_w} p_w h_i - \sum_{f \in F} \sum_{\pi \in \Pi} c_f^f z_f^f$$

$$\sum_{i \in I_w} h_i \leq \Delta_w^a = \sum_{\varsigma \in Z} P_{\text{air}}^{od,\varsigma} P_{\text{air}}^{od,\varsigma} \gamma_\varsigma^w d_w \quad \forall w \in W$$
The objective function \(12.6\) maximizes the airline’s profit. Constraints \(12.7\) state that the captured demand per market must be at most equal to the market share multiplied by the unconstrained demand in that market. Constraints \(12.8\) allocate the passenger demand to flights ensuring that the number of passengers on a flight must be at most equal to the airplanes’ total capacity multiplied by the maximum allowable load factor. Groups of constraints \(12.9\) and \(12.10\) are slot constraints for arrivals and departures, respectively. Constraints \(12.11\) and \(12.12\) ensure that the frequency in each \(od\) pair cannot be greater than the number of flights operated in that \(od\) pair by all fleet types. Constraints \(12.13\) state that the schedule must be symmetric, that is, the number of departures and arrivals in every airport must be same within the planning period. Constraints \(12.14\) are fleet
utilization constraints which ensure that the utilization of each fleet type must not be greater than the available one. Group of constraints 12.15 count the number of planes of fleet type $\pi$ on the ground in airport $g$ at the beginning of the planning period. Constraints 12.16 limit the number of grounded planes to the fleet size. Block of constraints 12.17 states a maximum number of planes to be grounded at each airport due to maintenance and crew issues. Constraints 12.18 introduce some regulations regarding flight frequencies pertaining to some flights that are subject to regulations and must be operated in any case. Groups of constraints 12.19-12.22 are variable value constraints.

### 12.4 Case Study and Results

We evaluate our model’s performance with case studies focusing on a single airline’s perspective. We use data provided by IBERIA, representing its operations for the year 2010. The dataset consists of operating schedule information, operating expenses, demand values, frequencies from other operators, and the available fleet. Air and rail competition has been considered. There are two main types of competing airlines: low cost and legacy airlines. There are three low cost airlines and three legacy airlines in this case study. The air-rail competition is present in six origin-destination pairs: Madrid-Barcelona, Barcelona-Madrid, Madrid-Sevilla, Sevilla-Madrid, Madrid-Málaga and Málaga-Madrid. There is one unique rail operator. Due to lack of data, we assume in this study case that there is no response from the competitors to the airline’s schedule changes suggested by our model. However, each operator will likely respond to each competitor’s schedule change in real life. Therefore, the airline under study will face a multi-operator and non-cooperative game. Each player (operator) will use alternative strategies to maximize its own pay-off function, the value of which depends on each of the players’ simultaneous actions. Hence, the game will consist of a set of operators with a set of strategies such as frequency of service, timetable, fleet size and ticket price. Accordingly, the airline under study will have to develop an alternative strategy every time a competitor changes its schedule, that is, it will solve the IASCOM to obtain a new strategy.

The air network is a pure hub-and-spoke network with 23 different airports. The hub is located in Madrid. There is no direct flight bypassing the hub airport. We discretize time at different airports at different levels based on the levels of operations by the carrier and the congestion levels of the airports. There are some airports in the Spanish network which have very low utilization (i.e., two flights per day). For these airports a discretization of two time periods per day is employed (one time period for each half-day). However, for
congested airports, such as airports in Madrid and Barcelona, a discretization of six time periods is implemented. There are 44 possible flights and 104 passenger routes within the network. There are three different fleet types available for this case study: an A-319 fleet with 141 seats per aircraft; an A-320 fleet with 171 seats per aircraft; and an A-321 fleet with 200 seats per aircraft. We have considered a planning period of seven days. We require the planning to be periodic, that is, the fleet distribution must be equal at the beginning and the end of the planning period.

The IASCOM is a non-linear mixed integer model. The non-linearity is in constraints \(12.7\). The captured demand is a non-linear function of the airline’s frequency values. In order to solve the model, we linearize this expression using piecewise functions. Consequently, we approximate the relationship between the fraction of passengers selecting the airline and the frequency values by a piecewise linear function. The piece sizes are selected to be the same for every origin-destination pair, equal to one frequency value. We coded the IASCOM in GAMS, using CPLEX 12 as the optimization solver, on a computer with 8 GB RAM and solved all models to a 1% relative gap. The computational time across all the test cases never exceeded 1043 seconds.

We present two different case studies in the following subsections. In Subsection 12.4.1 our base case scenario measures how closely our model solutions match reality, as described by IBERIA’s schedule. In Subsection 12.4.2 we introduce a new scenario involving the entry of high speed rail in a market, thus stimulating total demand volume in the market. We propose a model extension to capture this demand stimulation and then solve the extended IASCOM to the predicted response.

12.4.1 Base case Scenario

One of the main parameters in the IASCOM model is the maximum allowable load factor on every flight, designated by the load factor multiplier \(lfm_f\). Although there are historical data available on load factors, it is not obvious how to make use of it without forcing the IASCOM model solution to mirror IBERIA’s current schedule. Consequently, we solve IASCOM varying the maximum allowable load factor in order to compare the current schedule operated by IBERIA with the new schedules generated by the model. The aim of this sensitivity analysis is twofold: first, to validate our approach and second, to verify the robustness of the model formulation.

A base numerical value is needed for the load factor multiplier. We set this base value equal to the average load factor value \(alf_f\) for every flight leg \(f\), computed as follows:
1. If $f_{mf}$ is set to 1 for every flight.

2. The mathematical model in 12.6–12.22 is solved.

3. We look for the captured demand in each flight. If demand remains unserved, we set $alf_f$ equal to 1; otherwise it is set to 0.85.

Then, a sensitivity analysis is performed. Seven different experiment runs are conducted. The load factor multiplier is assigned a different value for each, with $lf_{mf} = \min\{\rho \cdot alf_f, 1\}$, where $\rho$ takes the values: 0.85, 0.90, 0.95, 1.00, 1.05, 1.10, 1.15.

We compare the solutions provided by the IASCOM with the solution operated by IBERIA. Table 12.2 shows a summary of different solutions for a simplified national network (it displays information for several OD pairs). These different solutions have been obtained varying the maximum allowable load factor per flight. Each column in Table 12.2 corresponds to a case where the load factor multiplier is different; its value is obtained as we have explained above. The last column in the table corresponds to the schedule operated by IBERIA. Each row shows the frequency value for the OD in the row. The row $TOTAL$ lists the total frequency value in the network, that is, the number of flights in the week. The row $ALF$ gives the average load factor obtained for each case.

Similar to Vaze and Barnhart [122], we use Mean Absolute Percentage Error ($MAPE$), with $MAPE_1$ (equation 12.23) measuring the error in the total frequency value in each origin-destination pair; $MAPE_2$ (equation 12.24) measuring the error in the frequency value per fleet type in each origin-destination pair; and $MAPE_3$ (equation 12.25) measuring the error in the total number of flown seats in each origin-destination pair.

$$MAPE_1 = \frac{\sum_{od \in OD} |\hat{f}_{od} - f_{od}|}{\sum_{od \in OD} f_{od}} \quad (12.23)$$

$$MAPE_2 = \frac{\sum_{f \in F} \sum_{\pi \in \Pi} |\hat{z}_{f}^{\pi} - z_{f}^{\pi}|}{\sum_{f \in F} \sum_{\pi \in \Pi} z_{f}^{\pi}} \quad (12.24)$$

$$MAPE_3 = \frac{\sum_{f \in F} \sum_{\pi \in \Pi} q_{\pi} \left(\hat{z}_{f}^{\pi} - z_{f}^{\pi}\right)}{\sum_{f \in F} \sum_{\pi \in \Pi} q_{\pi} z_{f}^{\pi}} \quad (12.25)$$

where $\hat{f}_{od}, \hat{z}_{f}^{\pi}$ are the solutions provided by the IASCOM optimization and $f_{od}, z_{f}^{\pi}$ are based on the current schedules operated by the airline.

Table 12.3 shows the values of the Mean Absolute Percentage Errors defined above. The IASCOM predicted frequency values are close to the current schedule operated by
Table 12.2 Comparing solutions while varying the load factor multiplier

<table>
<thead>
<tr>
<th>O.D</th>
<th>0.85</th>
<th>0.90</th>
<th>0.95</th>
<th>1.00</th>
<th>1.05</th>
<th>1.10</th>
<th>1.15</th>
<th>AIRLINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oviedo.Madrid</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Madrid.Oviedo</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Bilbao.Madrid</td>
<td>29</td>
<td>29</td>
<td>26</td>
<td>26</td>
<td>25</td>
<td>24</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>Madrid.Bilbao</td>
<td>29</td>
<td>29</td>
<td>26</td>
<td>26</td>
<td>25</td>
<td>24</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>Palma.Madrid</td>
<td>15</td>
<td>17</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>Madrid.Palma</td>
<td>15</td>
<td>17</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>Granada.Madrid</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Madrid.Granada</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>TOTAL</td>
<td>656</td>
<td>680</td>
<td>710</td>
<td>738</td>
<td>734</td>
<td>710</td>
<td>696</td>
<td>740</td>
</tr>
<tr>
<td>ALF</td>
<td>68.12</td>
<td>73.43</td>
<td>75.51</td>
<td>76.84</td>
<td>78.76</td>
<td>80.44</td>
<td>81.44</td>
<td>75.6</td>
</tr>
</tbody>
</table>

Table 12.3 Mean Absolute Percentage Errors

<table>
<thead>
<tr>
<th>ρ</th>
<th>MAPE_1 (%)</th>
<th>MAPE_2 (%)</th>
<th>MAPE_3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85</td>
<td>12.43</td>
<td>28.10</td>
<td>13.72</td>
</tr>
<tr>
<td>0.9</td>
<td>9.72</td>
<td>19.45</td>
<td>10.34</td>
</tr>
<tr>
<td>0.95</td>
<td>4.05</td>
<td>10.54</td>
<td>6.69</td>
</tr>
<tr>
<td>1</td>
<td>3.01</td>
<td>6.21</td>
<td>3.20</td>
</tr>
<tr>
<td>1.05</td>
<td>4.05</td>
<td>10</td>
<td>4.37</td>
</tr>
<tr>
<td>1.1</td>
<td>6.21</td>
<td>14.86</td>
<td>5.99</td>
</tr>
<tr>
<td>1.15</td>
<td>8.64</td>
<td>20.54</td>
<td>8.09</td>
</tr>
</tbody>
</table>

IBEIRIA. The error in fleet assignment is slightly higher than that in frequencies but is still small in an absolute sense. The variation in the number of seats flown is closely related to the variation in frequencies. In order to get deeper insights into the model’s predictions, Figures 12.1 and 12.2 represent the data used to calculate MAPE in a disaggregated way for ρ = 1. Figures 12.1 and 12.2 are drawn using a logarithmic scale with base 10. Figure 12.1 shows the actual frequencies per fleet type operated by the airline (on the x-axis) and the frequencies per fleet type given by the model (on the y-axis). Similarly, Figure 12.2 shows the actual number of seats flown by the airline (on the x-axis) and the number of seats flown by the airline as estimated by the model (on the y-axis).

Figure 12.3 shows the percentage difference in the air mode’s market share (that is, the percentage of passengers that all the airlines capture) compared to the actual values as provided in the data obtained from IBERIA, for the markets in which HSR currently operates. For the Madrid-Barcelona case, the predicted market share is never below the
Figure 12.1 Comparison of the frequencies per fleet type given by the model with the airline’s actual schedule for $\rho = 1$

Figure 12.2 Comparison of the seats flown of the model with the airline’s actual schedule for $\rho = 1$
actual value. This is due to the fact that in this market there is a constraint that imposes a minimum level of service and therefore, the frequency number is always above a minimum number (this minimum level is only active for $\rho = 0.85$ and $\rho = 1.15$). In the Madrid-Sevilla market, the variation in the air market share is highly dependent on the schedule predicted by the model. This is due to the fact that in this market the only operating airline is IBERIA; the model predicts that the airline should offer lower frequencies due to the competition of HSR. In the Madrid-Málaga case, the model predicts that the market share grows with the $\rho$ value.

Other important metrics of interest include the number of passengers transported and the operators’ total profit. Figure 12.4 shows the percentage difference in the number of transported passengers and in profit with respect to IBERIA’s actual values. Both behave in a similar way. It is interesting to note that when the difference in the number of passengers transported is close to zero, the difference in profit is also close to zero; when the difference is close to zero it is because the frequency and the number of passengers transported in every market is very close to those in the IBERIA schedule. This occurs when $\rho$ is around 0.98. Regarding connecting passengers, they increase as the maximum allowable capacity increases. Direct passengers are ‘more expensive’ than connecting passengers (when comparing yields). So, the model first serves first non-stop passengers.

The quality of the schedule determined by our model can be analyzed by considering the percentage variation in the average revenue per available seat kilometer (RASK), and the average revenue per passenger kilometer (RPK) with respect to IBERIA’s schedule. RASK and RPK are calculated as average values over the whole network. The RASK increases as $\rho$ increases; that is, as we increase the maximum allowable load factor, the aircraft can accommodate more passengers and the revenue per seat kilometer is increased. When the maximum allowable load factor is lower, the percentage difference becomes negative which means that airplanes have more empty seats. The RPK has the opposite behavior. When the number of captured passengers is low, the model solution serves the more expensive passengers, while when the maximum allowable load factor goes up, it serves more of the lower fare passengers. As we are using average fares, these variations happen because the model provide more frequency on higher fare routes with increasing load factor.

Fleet utilization depends on the schedule operated. Figure 12.5 shows the percentage difference in fleet utilization with respect to IBERIA’s actual values (fleet utilization is measured as the average block hours flown per day). The solid line corresponds to fleet type A-319, the dashed line to A-320s and the dotted line to A-321s. In general, fleet
Figure 12.3 Percentage difference (with respect to actual numbers) in the air market share in the markets where HSR currently operates.

Figure 12.4 Percentage difference in the number of transported passengers and in profit with respect to IBERIA’s actual values.
utilization is lower when the maximum allowable capacity in flights is low. This can be attributed to the fact that when the maximum allowable number of seats to be sold on each flight goes down, it becomes less profitable to operate certain flights. As a result, in some cases, it becomes more profitable to keep the aircraft on the ground rather than operate flights that are marginally profitable at higher values of maximum allowable load factor. As a result, the number of scheduled flights is also lower at lower values of maximum allowable load factors (see Table 12.2). Thus, the scheduled frequencies increase as the value of \( \rho \) increases, thereby increasing fleet utilization. Nevertheless, when the value of \( \rho \) becomes too high, scheduled frequencies decrease because the offered capacity in each flight is now so large that passenger demand is satisfied with fewer flights, thus decreasing fleet utilization, except for the A-321 fleet (which has the highest capacity). At higher values of \( \rho \), it becomes more economical to fly the largest planes. The cost per seat is lower because the planes are allowed to carry more passengers.

Figure 12.5 Percentage difference in the fleet utilization with respect to IBERIA’s actual values
12.4.2 Impacts of the Entry of High Speed Rail in a Market

In Spain, HSRs, mainly developed by the government are a strong source of competition to the air mode. They started operating in the origin-destination pairs Madrid-Sevilla and Sevilla-Madrid. Then, they expanded HSR operations to the other OD pairs including Madrid-Málaga, Málaga-Madrid, Madrid-Barcelona, Barcelona-Madrid, Madrid-Valencia and Valencia-Madrid. In all these cases, the introduction of HSRs meant a loss in the market share for the airlines. Consequently, the airlines had to change their way of operation by offering lower frequencies and smaller fleet sizes.

In the near future, the Spanish government is planning to operate HSRs from Madrid to the region of Galicia, in the northwest of Spain (Ministerio de Fomento [97]). Consequently, IBERIA will face the challenge again of modifying its schedule to compete not only against other airlines, but also against HSR.

We consider a scenario where HSR starts to compete simultaneously in six new origin-destination pairs, and we solve the IASCOM in order to study the predicted response. The origin-destination pairs affected by this new competition are those in the Galicia region: Madrid-Vigo, Vigo-Madrid, Madrid-La Coruña, La Coruña-Madrid, Madrid-Santiago and Santiago-Madrid. We assume that the remaining airlines do not change their schedules because we have no data on their likely schedule changes. All the experiments in this section are performed with $\rho = 1$ and using the available fleet as described for the base case scenario.

In the mathematical model presented in 12.6-12.22, we assume that the unconstrained demand is given and fixed for each market. This assumption is reasonable in many markets where the entry of new operators is unlikely. However, if entry occurs, the unconstrained demand will likely change due to demand stimulation. Consequently, we modify our model to account for this demand stimulation.

The objective is to explain as much of the variation in demand as possible, by identifying the attributes that have the greatest and most direct impact on the volume of OD market demand. The major factors affecting the volume of travel demand in an OD market are the price of travel, total trip time and demographics related to the market itself (Belobaba [15]). We will assume that the entry of HSR will change the average price of the trip in the market and that the frequency value will also vary. No significant demographic changes are expected to occur because we are looking at a planning period of several months. Consequently, a simple approach for modeling demand includes modeling unconstrained demand variation with the total frequency (as a measure of schedule displacement) and average price in the market. Thus, we assume unconstrained demand
to be elastic to frequency and price.

The concepts of price and time elasticity of demand for air travel can be incorporated into a relatively simple OD market demand function. Consider the following multiplicative model of demand for travel in a given market:

$$d_w = M_w \bar{p}_{od} \left[ \bar{t}t + \frac{sd}{f_{od}} \right]^\phi,$$

(12.26)

where $M_w$ is a market sizing parameter (constant), $\bar{p}_{od}$ is the average price of travel, $\bar{t}t$ is the average trip time, $f_{od}$ is the frequency value and $\epsilon$ and $\phi$ are price elasticity and time elasticity of demand, respectively. $sd$ is a constant expressed in hours. The schedule displacement component of total trip time may be expressed as $\frac{sd}{f_{od}}$. Values of $M_w$, $\epsilon$ and $\phi$ can be estimated from a historical data sample of $d_w$, $\bar{p}_{od}$ and $f_{od}$ for the same market, or from a sample of similar markets over a period of time. Statistical estimation techniques like ordinary least squares regression applied to historical data provide us with the best fit curve (Belobaba [15]).

We developed this model extension to be applied to the case when HSR enters as a new operator in a market where the airline was already operating. Consequently, we used a sample of similar markets over a period of time in order to estimate the values of $M_w$, $\epsilon$ and $\phi$. These similar markets are Madrid-Málaga, Madrid-Barcelona and Madrid-Sevilla, also affected by the entry of the HSR operator (the entry of the HSR is these markets was in years 2007, 2008 and 1992, respectively). Taking logarithms of both sides of equation [12.26] and then regressing using the ordinary least squares regression method, we have obtained the numerical values for the parameter $M_w$ and find values for $\epsilon$ and $\phi$ of -0.8476 and -1.1376, respectively. We use the classical F-test for linear models to test whether a naive model is better than the one in [12.26]. This test statistic equals 5,2031. Using the critical value of 3.4668 (2 degrees of freedom in the numerator and 21 degrees of freedom in the denominator; 0.95 confidence level), we reject the null hypothesis that the naive model is better than the model in [12.26].

Thus the unconstrained passenger demand in each market $w$ ($d_w$) is not assumed known apriori, instead it is assumed to vary as described in [12.26]. Therefore, the captured demand ($\Delta_w^a$) by airline $a$ in each market $w$ is $\Delta_w^a = \sum_{c} P_{a|air} P_{air|c} \gamma_{w|c} d_w$, where $d_w = M_w \bar{p}_{od} \left[ \bar{t}t + \frac{sd}{f_{od}} \right]^\phi$. Note that we are implicitly assuming here that the proportion of different types of passengers ($\gamma_{w|c}$) remains unchanged upon demand stimulation. The extended formulation is given by:
\[ \Delta w = \sum_{\varsigma \in Z} M_w \bar{p}_{\bar{w}} \left[ D + \frac{ad}{\sum_{a' \in A} \bar{p}_{a'} + \bar{p}_a} \right]^{\phi} \left[ e^{\log(a_{\bar{w}c}) + a_{\bar{w}f} \log(f_{\bar{w}d}) + b_{\bar{w}f} \log(p_{\bar{w}d}) + \gamma_{\bar{w}f} \log(m_{\bar{w}d})} \right] \right. \\
\left. \quad \cdot \left( e^{\theta_f \log \left( \sum_{a' \in A} e^{\log(a_{\bar{w}c}) + a_{\bar{w}f} \log(f_{\bar{w}d}) + b_{\bar{w}f} \log(p_{\bar{w}d}) + \gamma_{\bar{w}f} \log(m_{\bar{w}d})} \right) + \gamma_f \log(t_{\bar{w}f})} \right) \right] \right\}^{\bar{w}} \] 

Figure 12.6 shows the optimization model’s solution in terms of frequencies (on the y-axis) for several hypothetical scenarios characterized by different frequency values of operation of the HSR (on the x-axis). We assume a simultaneous entry of HSR into all six markets because they will be served along the same corridor. As before, the planning period is seven days. The average fare for HSR is assumed to be proportional to the fares in the current HSR markets. The response is similar for every market directly affected by the new competitor. As HSR starts operating in a new market, the airline’s frequency, as predicted by the optimization model, increases with increases in HSR frequency. As HSR frequency increases further, there is a point beyond which the predicted frequency of the airline starts decreasing until a threshold frequency value is reached by HSR. Beyond that point, the airline holds its frequency constant until another frequency threshold is reached by HSR. We assume connecting passengers are not affected by the entry of HSR, and they always choose to fly. However, fewer direct passengers select the airline due to competition from HSR until the flight is not profitable, and the optimal schedule for the airline does not serve those origin-destination pairs. Of course, the airlines can decide to continue operations in certain origin-destination pairs for reasons other than short-term profitability maximization, such as, maintaining the airline’s image or long-term growth strategy, etc.

In Figure 12.7, we depict the average model’s solution (for the airline) load factors for the origin-destination pairs Madrid-Vigo and Vigo-Madrid as a function of the scenario (characterized by the HSR frequency). The curves for the remaining origin-destination pairs with HSR entry are similar. Average load factor decreases with increasing schedule frequencies as the airline attempts to capture more passengers (see Figure 12.6). When the airline’s frequency is decreased, the average load factor sharply increases. However, it starts decreasing again until the point when flights are not profitable and service is discontinued.

With our optimization model, we study tactical competition in which ticket price is assumed to have an average and constant value. However, the entrance of new competitors...
Figure 12.6 Model’s response to HSR in origin-destination pairs Madrid-Vigo, Madrid-La Coruña and Madrid-Santiago

Figure 12.7 Average load factor predicted by the model in response to HSR entry with varying frequencies in the origin-destination pairs: Madrid-Vigo and Vigo-Madrid
is likely to cause changes in these average values. We analyze what might happen under each of four different price reduction scenarios, corresponding to a 10%, 20%, 30% and 40% reduction in average prices in markets where the new competitor has entered. We compare these four scenarios with the base scenario (with original prices). The reduction in average prices is only applied to the tickets of the airline under study, that is, IBERIA.

In Figure [12.8] we show the market share predicted by our model in response to HSR entry in the origin-destination pairs Madrid-La Coruña and La Coruña-Madrid for the base price and for 10%, 20%, 30% and 40% reductions in base price. The curves for the remaining pairs with HSR entry are similar. As shown, as the ticket price is lowered from 0% through 20%, market share rises (apart from the initial drop). By reducing average ticket price, our model predicts that IBERIA could compete for a larger set of scenarios against HSR. However, this is not true for ticket price reductions from 20% to 40%. By offering greater discounts in average prices, flights start becoming non-profitable. Our model results show that the number of frequencies should be lowered and therefore, market share drops. It is clear that for any scenario involving HSR entry, the loss of market share for IBERIA would be significant.

![Model’s predicted market share in response to HSR in origin-destination pair Madrid-La Coruña and La Coruña-Madrid](image)

**Figure 12.8** Model’s predicted market share in response to HSR in origin-destination pair Madrid-La Coruña and La Coruña-Madrid

The overall impact of this new scenario on total profit is shown in Figure [12.9] Total
profit prediction reaches a constant value when HSR frequency ceases to have any impact on our model’s predicted profit. This is due to the fact that our model predicts that IBERIA should exit from those markets where HSR has entered. Thus, there comes a point where increments in HSR’s frequency value have no effect on IBERIA’s profit. For the scenarios with a proposed moderate ticket price reduction (10% and 20%), the total profit predicted by our model is greater than the profit in the base scenario. However, aggressive discounts in the average prices do not result in further increases in profit because the operation of flights on those discounted markets might not be profitable due to the decreasing revenue per passenger. Moreover, due to these discounts, other markets become ‘more profitable’ (compared to the discounted markets) and the optimized schedule increases the number of flights in those markets.

![Figure 12.9 Model’s predicted total profit variation in response to HSR entry for the base price and reductions of 10%, 20%, 30% and 40% in ticket prices](image)

These experiments have been performed assuming a fleet of size and composition matching that of IBERIA. The results suggest that it might be advisable for IBERIA to make changes in its fleet to compete in markets with HSR entry. In the absence of HSR competition, the optimized schedule utilizes a heterogeneous fleet in the markets in the Galicia region (see Table 12.4). However, when HSR starts to operate, the only aircraft utilized in the optimized schedule are those with the smallest number of seats.
This effect is also observed in the scenarios with price reductions. However, the price reductions slightly change the fleet utilized. For the scenarios with low HSR frequency, the optimized schedule utilizes a heterogeneous fleet. Nevertheless, as HSR competes with higher frequencies, the optimized fleet again uses the smallest aircraft size exclusively. This suggests that the availability of smaller aircraft would make it possible for the airline carrier to compete more effectively against HSR.

Table 12.4 shows different scenarios’ solutions for OD pairs La Coruña-Madrid and Madrid-La Coruña (the results for the rest of OD pairs present a similar behavior). Different scenarios correspond to different HSR frequency values. Each column in Table 12.4 corresponds to a case where the HSR frequency \( f_{rail}^{od} \) is different. Each row shows the model predicted frequency value for the OD in the row. The frequency value per fleet type is displayed. For example, 0/1/10 implies that there are 0 flights per week with the A-319 fleet type, 1 with the A-320 and 10 with the A-321 fleet type. The second column in the table assumes a scenario where no HSR is competing. The third column assumes that the HSR competes with a frequency of seven and so on. In order to show what happens under each of the four different price reduction scenarios (corresponding to a 10%, 20%, 30% and 40% reduction in average prices in markets where the new competitor has entered), we show the frequency value per fleet type for each of them. Then, the rows are divided in 5 different sub-rows (only in the columns where the HSR competes): the first sub-row shows the base price scenario solution, the second sub-row shows the 10% reduction price scenario solution, the third sub-row the 20% price reduction scenario solution, the fourth sub-row the 30% price reduction scenario and the fifth sub-row the 40% price reduction scenario solution.

12.5 Summary

We have proposed a mixed integer and non-linear programming model for the schedule design and fleet assignment problem that includes a passenger choice model to capture multi-modal competition between high-speed rail, low-cost airlines, and legacy airlines. Customers’ behavior is modeled with a nested logit model of two stages: in the first stage the customers choose the mode and in the second stage the airline. The nested logit model has been calibrated using real data.

The optimization model has been tested using realistic problem instances obtained from the network of IBERIA. We have shown how the model is able to replicate IBERIA’s current decisions. We have evaluated multiple scenarios involving entry of High Speed
### Table 12.4 Comparing solutions for different scenarios for OD pairs La Coruña-Madrid and Madrid-La Coruña

<table>
<thead>
<tr>
<th>O.D</th>
<th>f_{rail}</th>
<th>0</th>
<th>7</th>
<th>14</th>
<th>21</th>
<th>28</th>
<th>35</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>La Coruña-Madrid</td>
<td>6/6/0</td>
<td>18/0/0</td>
<td>13/1/1</td>
<td>15/0/0</td>
<td>16/0/0</td>
<td>8/0/0</td>
<td>0/0/0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7/7/5</td>
<td>7/7/3</td>
<td>7/7/0</td>
<td>7/5/0</td>
<td>10/2/0</td>
<td>12/0/0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7/7/5</td>
<td>5/7/4</td>
<td>5/7/2</td>
<td>8/4/0</td>
<td>10/2/0</td>
<td>12/0/0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10/5/0</td>
<td>7/3/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td></td>
</tr>
<tr>
<td>Madrid-La Coruña</td>
<td>6/6/0</td>
<td>18/0/0</td>
<td>13/1/1</td>
<td>15/0/0</td>
<td>16/0/0</td>
<td>8/0/0</td>
<td>0/0/0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7/7/5</td>
<td>7/7/3</td>
<td>7/7/0</td>
<td>7/5/0</td>
<td>10/2/0</td>
<td>12/0/0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7/7/5</td>
<td>5/7/4</td>
<td>5/7/2</td>
<td>8/4/0</td>
<td>10/2/0</td>
<td>12/0/0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10/5/0</td>
<td>7/3/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>0/0/0</td>
<td></td>
</tr>
</tbody>
</table>

Rail and accounting for demand stimulation as a result of the new services.
Part IV

CONCLUSIONS, FUTURE RESEARCH & BIBLIOGRAPHY
Chapter 13

THESIS CONCLUSIONS

In the following, we describe the conclusions of this thesis. We detail the conclusions of each chapter in a separate way.

13.1 Chapter 5: Robust Rolling Stock in Rapid Transit Networks

The rolling stock model presented in this chapter is a new approach in the emerging area of urban Rapid Transit network optimization. The results are satisfactory, because in addition to commercial train services, they also account for empty movements, adequate allocation of material in the depots, the optimal train unit mix forming the trains, and rotation times. Another aspect considered in the model is the simultaneous inclusion of all daily rush hours.

Robustness is introduced into the model through a number of different approaches. First, composition changes are penalized depending on the depot stations and time periods. At congested depot stations, it might be very difficult to perform a composition change. Similarly, empty movements are penalized because they use the same infrastructure as commercial train services.

The results obtained in the network tests are satisfactory: operating costs are lowered while a high level of service quality for passengers is maintained and robust plans for network operation are provided. Moreover, the time needed to obtain these plans is reduced from the current system of manual planning under great time pressure. In addition, the possibility of analyzing several scenarios rather than just one is considered quite useful. For example, by varying penalties for excess passengers, we can obtain different solutions
providing different qualities of passenger service (i.e., the quality of service is parametrized using the mentioned penalties).

Although we have solved case studies in a modular way, as is done in real life, we have checked that the model sizes obtained are similar to, or even greater than, those in the related literature. In addition, we show how the computational times are sufficiently low to apply this approach in real life.

13.2 Chapter 6: Robust Routing of Rapid Transit Rolling Stock

The robust rapid transit routing model is a first approach to the new subject of rapid transit urban network logistics, which has been planned manually until now. The routing is studied with commercial movement, empty movement and adequate allocation of the train units in the depots. The data are obtained from a rolling stock model, in which the optimal combinations and splitting of the train units to form the trains are also included.

The results obtained in the network tests are satisfactory, not only because of the robustness of the sequences of each specific train unit, but also because they are obtained near real time. Furthermore, routing solution complements the rolling stock output because the last one cannot be realized without the first one.

We also include aspects from crew planning, i.e., we minimize the crew resources necessary to perform all duties. However, further research is needed to generate complete crew planning.

Some computational experiments using the RENFE network are presented. Line C5, with the highest frequencies in the network, is solved in reasonable computational times. In addition, two joint lines, C3 and C4, are also solved. In both cases, the expected delay is significantly reduced. This reduction implies that expensive swapping operations are also reduced.

13.3 Chapter 7: Integration of Timetable Planning and Rolling Stock in Rapid Transit Networks

In this chapter an integration of the railway timetabling and rolling stock assignment problem is formulated in the emerging area of urban rapid transit networks. The planned frequencies are known from the railway line planning problem. However, the proposed
model updates frequencies providing a more efficient network operation.

Services constraints are included ensuring that frequencies for some arcs are maintained in a determined frequency window. Moreover, these constraints also ensure that the headway is maintained for every train line in the network. In the passenger constraints for the demand, passengers in excess and unattended passengers are considered; we include different capacities depending on standing passengers’ density. Finally, in the rolling stock constraints shunting operations, fleet and depot capacities, and symmetry constraints are included.

In the model operating and leasing costs are minimized jointly with penalties for robustness. Robustness is defined through penalizing difficult shunting operations. Another robust approach is also included: ensuring parkings in strategic depot stations in order to perform swapping operations. Maintenance costs are included adding to the objective function maintenance daily costs for each train unit.

RENFE data have been used for the computational tests for the rapid transit network in Madrid. The presented service network design model considers the possible train lines in that network. The possible train lines are defined as those services between depot stations departing during each time period. Among them, the model chooses the optimal services (i.e.: the number of scheduled services are about 370 and 410 for Lines C5 and C3&C4, respectively). Computational results show how the rolling stock assignment depends on the efficiency of the railway timetable. We do not only change the timetable; also updating frequency values we introduce a greater freedom. That is, comparing with the rolling stock assignment with fixed timetable, we can see how better solutions can be obtained improving the overall solution efficiency and getting a greater robustness degree. We have also studied the possibility of minimizing maintenance costs for every train unit. However, using this approach, solutions with a lower degree of robustness are obtained because more composition changes are needed, which deteriorates the network operation.

The proposed model treats passenger demand from a centralized point of view. We represent passenger demand as a passenger flow through arcs. In order to decide the new timetable, passenger flow is aggregated in time, that is, every flow through a determined arc during a certain time interval is aggregated into the same time period. We have tried different time intervals’ lengths to aggregate passenger flows showing that it must be longer than trains’ headway. However, as this time length is increased, passengers’ behavior is worsen in the model formulation, so a tradeoff time length must be selected.
13.4 Chapter 8: Improving Robustness of Rolling Stock Circulations in Rapid Transit Networks

We have proposed a new approach to solve the rolling stock circulations. In order to obtain them two different problems must be addressed: the rolling stock assignment problem and the train routing problem. Up to now, they have been solved in a sequential manner, that is, the solution of one of them is known before solving the other one. Our approach proposes to integrate both problems. Consequently, it provides the global optimum to the problem.

The integrated approach is a good frame to improve the robustness degree of the system. We get the robustness through different points of view: rolling stock operations and operations’ connections. In rolling stock operations we may cite empty trains and composition changes. They are always difficult operations and complicate the network operation. A way of introducing robustness here is penalizing selectively these operations in order to ameliorate their possible negative effects in the network operation. The other way of introducing robustness we have presented is related to schedule performance and punctuality. In order to perform operations’ connections properly, the schedule must be match. However, the schedule is not always operated as planned and deviations from the planned operations may produce delays in operations. We account for these possible delays in order to produce a resistant schedule to them.

The proposed integrated model to solve the rolling stock circulations has an enormous size to be solved by the current commercial software. Therefore, we propose to decompose it using Benders decomposition. Using this technique the submodel may be reformulated in order to make it easier to be solved. However, for computational reasons we propose a Benders based heuristic to solve the proposed model: Benders based cuts are generated in each iteration in order to improve the rolling stock circulations.

Computational experiments show how the current solution operated by RENFE can be improved: more robust and smoother solutions are obtained. RENFE planners do not use operations research techniques for planning purposes. Therefore, the planning is greatly improved. However, the proposed integrated approach also improves the solution obtained using operations research techniques in a sequential manner. We are able to produce plans with fewer composition changes which are considered to be dangerous by planners. Selectively penalizing them we ameliorate their probability to produce negative and cascading effects in the network operation. Moreover, we also account for delays in the network operation. We have reduced the number of operations’ connections with
positive propagated delay. Consequently, the number of re-scheduling operations (i.e., swapping operations) are reduced making easier to the operator to recover.

The solution provided by our integrated approach is the schedule to be implemented by the operator. Consequently, the available planning horizon is enough to consider the needed computational times reasonable for implementation. Nevertheless, development of more intelligent heuristics to reduce the model size and computation costs may be defined.

13.5 Chapter 9: Recovery of Disruptions in Rapid Transit Networks

In this chapter we study the recovery problem of rapid transit networks. When dealing with a disruption, the operator wants to offer a service quality as good as possible while the system is being recovered to the original planning.

We propose a two-step approach to adjust the timetable and the rolling stock assignment, and we explicitly take the passengers’ reaction to the disruption into account. Our approach first computes the anticipated passenger demand. Then the integrated timetabling and rolling stock scheduling problem as a Mixed Integer Linear Programming model. Further, we embed the two-step approach in an iterative framework: demand is re-computed in every iteration using the schedule from the previous iteration. Therefore, the iterative process allows us to study demand response to each schedule.

In computational tests on realistic instances of RENFE, our method is able to find solutions with a very good balance between the managerial goals. Preliminary discussions with practitioners revealed that the solutions captured all important real-life restrictions, and have a good chance to be implementable in practice. Our computational times amount to a few minutes which is sufficiently close to the needs of real-time decision making. This is a great advantage with respect to the current system of manual re-planning where planners work under great time pressure.
13.6 Chapter 11: Robust Passenger Oriented Timetable and Fleet Assignment Integration in Airline Planning

The interaction between supply and demand has led us to propose a new robust approach to solve the airline scheduling problem, where schedule design, fleet assignment and passenger use problems are jointly solved.

Connecting passengers must perform an intermediate stop. Consequently, they need some time, which may vary, to accomplish it. This undetermined time is captured through a statistical distribution in order to introduce robustness into the model accounting for expected misconnected passengers. Therefore, the expected costs that the operator would incur due to misconnected passengers are reduced.

The model has been tested in a IBERIA’s simplified network. Computational results show how robustness is achieved. However, this robustness has a price. The robust approach has been compared with a non-robust approach showing the price of the achieved robustness. The price of robustness may remain in different issues: if we want to maintain the fleet utilization high in the robust solution, the price of robustness will remain in passengers: fewer passengers will be attended in the robust schedule. However, if we allow the fleet utilization to get lower, more passengers will be attended and the price of robustness will be shared between both the operator and passengers; the operator will have a schedule with a lower fleet utilization what may mean that it will be harder to amortize fleet resources and more crews may be needed to operate the robust schedule. Anyway, misconnected passengers are reduced in both solutions, achieving the robustness we were looking for.

13.7 Chapter 12: Integrated Airline Scheduling: Considering Competition Effects and the Entry of the High Speed Rail

The airline planning process involves solving problems such as frequency planning, timetable planning and fleet assignment. All of them are directly related to passenger demand: it influences in the schedule and the schedule influences in passenger demand. Consequently, appropriate customer behavior modeling is needed in order to be able to satisfy them.
Moreover, competition from different modes and operators make it difficult to estimate the passengers that an airline will capture.

We develop a tactical competition model for an airline-considering multi-modal competition between air and high-speed rail, and airline competition between legacy and low cost carriers-using a nested logit model of demand behavior. We use it to develop an integrated optimization schedule design model that includes frequency planning, approximate timetable development, fleet assignment and passenger demand choice. The model accounts for passenger demand competition, and captures the impacts of schedule decisions on passenger demand, as suggested by the S-curve relationship.

We calibrate the nested logit model using real data provided by IBERIA (the major Spanish airline). Then, we test the integrated optimization model in a real network from IBERIA including other air and rail transportation options in Spain. We conduct sensitivity analysis on various parameter values and demonstrate how the model is able to reproduce IBERIA’s current decisions. We evaluate multiple scenarios involving entry of High Speed Rail in some markets, and we account for the possibility of demand stimulation as a result of the new services. Depending on the High Speed Rail’s offered attributes, the model predicts what IBERIA should do in order to retain passengers if possible.

In summary, we show that our integrated optimization model is able to replicate current IBERIA’s decisions and, therefore, it may be used to predict its response to market changes such as service modifications by current competitors and entry of new operators. Hence, the proposed modeling approach is attractive from the perspective of the operator.
Chapter 14

FUTURE RESEARCH

This chapter briefly describes future research topics in railway and airline applications.

14.1 Railway Applications

All the research in railway applications we have presented in this dissertation is limited to deterministic demand data. However, in actual operations passenger demand usually varies on a daily basis, so that projected demand may not reflect the actual daily passenger demand, where stochastic disturbances may occur. Passenger demand fluctuations arising from stochastic demands could affect the actual performance of the planned schedules. In other words stochastic disturbances arising from variations in daily passenger demand could affect the optimality of the rolling stock schedules and timetables. Therefore, to set a good schedule, not only supply has to be considered, but passenger demand fluctuations arising from stochastic demands in actual operations also have to be taken into account.

Apart from stochastic demands, market share may also vary with passenger choice behaviors, especially where competitive forces are present. Therefore, demand choice modeling is a very interesting topic. It is crucial to properly predict the amount of passengers a certain operator is going to capture (i.e., the actual market share may decrease with respect to the projected market share if a schedule is inferior, and vice versa). An important issue in competition is to investigate game-theoretic models of competition and derive results that provide insights into important characteristics of equilibrium.

Demand choice modeling is also very interesting for scheduling in competitive markets. Nowadays, the European Union is encouraging competition among different railway operators. In Spain, the entry of new rail operators is expected to occur soon. Consequently, RENFE will have to face competition form different operators. New schedules will be
needed to compete and to be able to survive.

An ambitious and difficult topic for future research would be the integration of all phases of the planning process (i.e., the planning of timetables, rolling stock assignment, routing and crew scheduling). Such an ideal system, of which we should not lose sight, is the final goal and would take into account every aspect and level in the planning process.

In the field of recovery from disruptions, we are going to embark on a deeper study of Pareto optimal solutions in order to give the operator a wider and deeper insight in the different recovery solutions depending on the objective function terms weights. Further research needs to refine the multinomial logit model in order to better capture the passengers’ behavior. For this purpose more real-life data must be gathered and studied.

The railway operator is interested in developing recovery plans that recover the whole system as fast as possible. The quality of a recovery plan may be measured by the number of changes we have to perform in order to recover to the original schedule once the disruption has ended. Therefore, controlling the recovery length is a very interesting topic for future research.

Another quite challenging option would be to develop an optimization model to deal with the complex interaction of timetable, rolling stock and passengers. This option would need an integrated optimization model dealing with origin-destination demand. Further, the integrated recovery of the previous problems and crews is a very interesting and challenging problem.

Apart from developing recovery plans, robust plans must be implemented. We have studied several problems in the railway planning process where robustness has been introduced. However, it may occur that robust plans are very expensive to be operated in a regular daily basis. In such a case, recoverable robustness is of high interest. Recoverable robustness is a methodology for producing solutions that are able to be recovered from a set of disrupted scenarios, using a restricted number of recovery algorithms in a limited effort. Consequently, we need to define a recovery algorithm in order to recover from each different disruption or scenario.

14.2 Airline Applications

It is widely accepted in the airline industry that the demand is uncertain. This uncertainty usually makes demand to vary on a daily basis, so that projected demand may not reflect the actual daily passenger demand, where stochastic disturbances may occur. Passenger
demand fluctuations could affect the actual performance of the planned flight schedules. In practice, the performance of an optimal plan could be reduced when applied to actual operations where passenger demand fluctuations occur. Therefore, to set a good flight schedule, not only does the fleet and related supply have to be considered, but passenger demand fluctuations arising from stochastic market demands and variable market share in actual operations also have to be taken into account.

We have presented as a contribution in this dissertation the integration of different problems in the airline planning process. Although practical, the sequential nature of schedule optimization leads to suboptimal plans, with potentially economic losses. Improved plans can be generated by building and solving integrated models. An ambitious topic for future research would be the integration of more phases of the planning process.

Crew cost is the second largest operating expense. Therefore, it is crucial for the airline efficient crew schedules. Integration of crew scheduling with prior planning problems such as passenger demand use, fleet assignment and maintenance routing and is of high interest.

It is also interesting to integrate scheduling models, which minimize operating costs, with competition models, which maximize revenues. We have presented a novel approach which provides airline schedules accounting for competition at the tactical level. A natural extension of this work is to provide a tool for adjusting these tactical schedules by determining flight timetables and balancing fleet resources in the network. A key area of research is the estimation of parameters for passenger choice-based schedule development. This is a challenging task because it usually involves censored data form competing operators.

Although we have presented an approach for airline scheduling considering competitive effects, we have not considered the game-theoretic aspects of such competition. A critical aspect in airline competition is to investigate game-theoretic models of airline competition and derive theoretical results that provide insights into important characteristics of equilibrium. However, this is a challenging task because of the number and different type of operators involved in the game.

A new development that improves the schedule efficiency and profitability is the implementation of demand driven dispatch where fleet assignment changes are made between crew compatible aircraft close to the day of departure to better match passenger demand. The integration of this approach with passenger choice models is a promising area of research.

The schedule planning and optimization processes at airlines produce plans that are
rarely execute. During operation of the schedule, disruptions result in the need to replan and create recovery plans. Creating these responses is particularly challenging because of the need to develop feasible plans in a very short time period. Much of the work to date focuses on the recovery of a particular resource. However, disruptions always affect multiple resources. Therefore, an interesting future research topic is to develop integrated recovery plans. Airlines are currently seeking for an integrated recovery model which deals with passenger, schedule and crew recovery plans in matter of minutes; this is a very challenging future research because of the problem size and the need of solving it quickly.

Developing robust plans has also attracted much of the research in this area. Robust plans are usually easier to repair when replanning is necessary. However, they are usually expensive to be operated in a regular daily basis. Therefore, developing recoverable robust plans will be a challenging topic of research.
Bibliography


LUIS CADARSO MORGA

Education

  
  Dissertation on solving large-scale mathematical optimization models applied to the aviation and railway industries.

• Master Science, Aerospace Engineering, June 2011.

• Aerospace Engineering, June 2009.

Experience

• Technical University of Madrid
  
  - Research Assistant, 2007-Present

  Presented research at leading conferences including the Institute for Operations Research and the Management Sciences (INFORMS), European Conference on Operational Research (EURO) and the International Federation of Operational Research Societies (IFORS).

  Original research disseminated through five publications in international top research journals.

  Collaborated with decisions makers at RENFE (the major train operator in Spain): timetable planning, rolling stock assignment, train routing, robustness, recoverability from disruptions, etc.

  Collaborated with decisions makers at IBERIA (the major airline in Spain): scheduling, fleet assignment, aircraft routing, robustness, competition against airlines and high speed rail, demand modeling, etc.

  - Teaching Assistant, 2010-2012
Prepared and delivered lectures on topics of optimization (airline and railway optimization).

- Massachusetts Institute of Technology

- Rotterdam School of Management, Erasmus University

**Awards**


**Papers**

[1] Authors: Luis Cadarso, Ángel Marin
  Title: Robust routing of rapid transit rolling stock
  Journal: Public Transport 2(1-2):51-68

[2] Authors: Luis Cadarso, Ángel Marin
  Title: Robust rolling stock in rapid transit networks

[3] Authors: Luis Cadarso, Ángel Marín
  Title: Integrated Robust Airline Schedule Development
  Journal: Procedia Social and Behavioral Sciences 20:1041-1050

[4] Authors: Luis Cadarso, Ángel Marín
  Title: Integration of timetable planning and rolling stock in rapid transit networks

[5] Authors: Luis Cadarso, Ángel Marín, Gábor Marótı
  Title: Recovery of Disruptions in Rapid Transit Networks
[6] Authors: Luis Cadarso, Ángel Marín
Title: Robust Passenger Oriented Timetable and Fleet Assignment Integration in Airline Planning
Journal: Journal of Air Transport Management 26:44-49

[7] Authors: Luis Cadarso, Ángel Marín
Title: Recoverable Robustness in Rapid Transit Network Design
Journal: Procedia - Social and Behavioral Sciences 54:1288-1297

Conferences

[1] Authors: Ángel Marín, Luis Cadarso
Title: An approach to Rapid Transit Rolling Stock Assignment
Conference: 3th International Seminar on Railway Operations Modelling and Analysis
Place: Zurich, Switzerland Date: February 2009

[2] Authors: Ángel Marín, Luís Cadarso
Title: The Rapid Transit Routing Problem
Conference: CASPT’2009
Place: Hong Kong Date: July 2009

[3] Authors: Ángel L. Matesanz, Isaac Musat, Ángel Marín, Luis Cadarso
Title: Inferencia Estadística en el Análisis de los Retrasos en una Red Ferroviaria
Conference: 2ª Jornada sobre Modelos de Optimización Aplicados a la Planificación Robusta y la Gestión de los Servicios de transporte Público en Caso de Emergencia (More 2009)
Place: Sevilla, Spain Date: November 2009

[4] Authors: Ángel Marín, Luis Cadarso
Title: Asignamiento óptimo del Material Rodante en Líneas de Cercanías
Conference: 2ª Jornada sobre Modelos de Optimización Aplicados a la Planificación Robusta y la Gestión de los Servicios de transporte Público en Caso de Emergencia (More 2009)
Place: Sevilla, Spain Date: November 2009
[5] Authors: Ángel Marín, Luis Cadarso
Title: Reasignamineto Robusto del Material Rodante en Redes de Cercanías
Conference: 2ª Jornada sobre Modelos de Optimización Aplicados a la Planificación Robusta y la Gestión de los Servicios de transporte Público en Caso de Emergencia (More 2009)
Place: Sevilla, Spain Date: November 2009

[6] Authors: Luis Cadarso, Ángel Marín
Title: Asignación de Horarios y Material Rodante
Conference: Workshop on Locational Analysis and Related Problems 2010
Place: Sevilla, Spain Date: February 2010.

[7] Authors: Luis Cadarso, Ángel Marín
Title: Robust Airline Logistics
Conference: Optimization Days 2010
Place: Montreal, Canada Date: May, 2010

[8] Authors: Ángel Marín, Luis Cadarso
Title: Rapid Transit Networks: Integrating Logistic Planning
Conference: ALIO-INFORMS Joint International Meeting
Place: Buenos Aires, Argentina Date: June 2010

[9] Authors: Luis Cadarso, Ángel Marín
Title: Integrated Airline Robust Scheduling and Fleet Assignment
Conference: ALIO-INFORMS Joint International Meeting
Place: Buenos Aires, Argentina Date: June 2010

[10] Authors: Ángel Luis Matesanz, Isaac Musat, Luís Cadarso, Ángel Marín
Title: Retrasos en la Red de Renfe Cercanías Madrid
Conference: IX Conference de Ingeniería de Transporte 2010
Place: Madrid, Spain Date: July 2010

Title: Reasignamiento Robusto del Material Rodante
Conference: IX Conference de Ingeniería de Transporte 2010
Place: Madrid, Spain Date: July 2010

[12] Authors: Luis Cadarso, Ángel Marín, Gabor Maróti, Lars Nielsen
Title: Integrating Timetable Recovery and Rolling Stock Rescheduling  
Conference: WCRT’2010  
Place: Lisboa, Portugal Date: July 2010

[13] Authors: Ángel Marín, Luis Cadarso  
Title: Integrated Time Table and Rolling Stock  
Conference: EURO’2010  
Place: Lisboa, Portugal Date: July 2010

[14] Authors: Luis Cadarso, Ángel Marín  
Title: Robust Airline Scheduling  
Conference: PANAM  
Place: Lisboa, Portugal Date: July 2010

[15] Authors: Luis Cadarso, Ángel Marín  
Title: Passenger Oriented Airline Scheduling and Fleet Assignment  
Conference: INFORMS 2010  
Place: Austin, Texas Date: November 2010

[16] Authors: Luis Cadarso, Ángel Marín  
Title: Integración Robusta de la Logística Ferroviaria: Horarios y Material Rodante.  
Conference: 3ª Jornada sobre Modelos de Optimización Aplicados a la Planificación Robusta y la Gestión de los Servicios de transporte Público en Caso de Emergencia (More 2010)  
Place: Madrid, Spain Date: November 2010

[17] Authors: Luis Cadarso, Ángel Marín  
Title: Integración Robusta de la Logística Ferroviaria: Material Rodante y Enrutamiento.  
Conference: 3ª Jornada sobre Modelos de Optimización Aplicados a la Planificación Robusta y la Gestión de los Servicios de transporte Público en Caso de Emergencia (More 2010)  
Place: Madrid, Spain Date: November 2010

[18] Authors: Luis Cadarso, Ángel Marín  
Title: Airline Scheduling Under Competition  
Conference: Exploratory Workshop on MINLP.  
Place: Sevilla, Spain Date: December, 2010
[19] Authors: Luis Cadarso, Ángel Marín  
Title: Airline Planning Considering Competition Effects  
Conference: Workshop on Locational Analysis and Related Problems 2011  
Place: Las Palmas de Gran Canaria, Spain Date: February 2011.

[20] Authors: Luis Cadarso, Ángel Marín  
Title: Robust rolling stock and routing integration in rapid transit networks  
Conference: RAILROME 2011  
Place: Roma Date: February 2011.

[21] Authors: Luis Cadarso, Ángel Marín, Gábor Maróti  
Title: Recovery of Disruptions in Rapid Transit Networks  
Conference: IFORS 2011  
Place: Melbourne, Australia Date: July 2011.

[22] Authors: Luis Cadarso, Ángel Marín  
Title: Introducing Competition in Airline Schedule Development  
Conference: IFORS 2011  
Place: Melbourne, Australia Date: July 2011.

[23] Authors: Luis Cadarso, Ángel Marín  
Title: Integrated Robust Airline Schedule Development  
Conference: EWGT 2011  
Place: Poznan, Poland Date: September 2011

[24] Authors: Luis Cadarso, Ángel Marín  
Title: Competition and Pricing in Airline Scheduling  
Conference: INFORMS 2011  
Place: Charlotte, USA Date: November 2011

[25] Authors: Ángel Marín, Luis Cadarso  
Title: Robustez recuperable en problemas de diseño y logística de redes de transporte  
Conference: SEIO 2012  
Place: Madrid, Spain Date: April 2012

[26] Authors: Luis Cadarso, Ángel Marín, Gábor Maróti  
Title: Recovery from Disruptions with Recovery Length Controlled  
Conference: EURO 2012
[27] Authors: Luis Cadarso, Ángel Marin
Title: Recoverable Robustness in Rapid Transit Network Design
Conference: EWGT 2012
Place: Paris, France Date: September 2012

[28] Authors: Luis Cadarso, Cynthia Barnhart, Virot Chirapandakul, Ángel Marin, Vikrant Vaze
Title: Airline Scheduling in Competitive Markets
Conference: INFORMS 2012
Place: Phoenix, Arizona Date: October 2012