Air-rail timetable synchronisation : passenger demand estimation, optimisation models, and application to the Western Europe case study

Clara Buire

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

Université de Toulouse

Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

Synchronisation des plannings train-avion : estimation de la demande passager, modèles d'optimisation mathématique, et application au cas de l'Europe de l'Ouest

---

Thèse présentée et soutenue, le 18 avril 2024 par

Clara BUIRE

École doctorale
AA - Aéronautique, Astronautique

Spécialité
Mathématiques et Applications

Unité de recherche
ENAC-LAB - Laboratoire de Recherche ENAC

Thèse dirigée par
Daniel DELAHAYE et Aude MARZUOLI

Composition du jury
M. Cyrille BRIAND, Président, Université Toulouse III - Paul Sabatier
M. Andrew COOK, Rapporteur, University of Westminster
Mme Sophie DEMASSEY, Rapporteure, PSL Université de Paris
M. Denis MONTAUT, Examinateur, Eurodecision
M. Daniel DELAHAYE, Directeur de thèse, ENAC
Mme Aude MARZUOLI, Co-directrice de thèse du monde socio-économique, OnSolve

Membres invités
M. Marcel Mongeau, ENAC

Air-rail timetable synchronisation :
Passenger demand estimation, optimisation models, and application to the Western Europe case study

Abstract

Air transportation network has been developed and optimised for years in order to provide passengers with a high level of service. However, the growing increase in environmental awareness and airport congestion make trains a relevant alternative to replace short-haul flights. If trains have to complement flights, collaboration between both operators will be required to maintain attractiveness and limit passengers' discomfort during transfers between modes. In this thesis, we propose synchronising air and rail timetables to enhance the quality of air-rail transfers for passengers at hub airports. As the transfer demand between rail and air is not publicly available, we first introduce a methodology to generate realistic air-rail transferring passenger demand using open-source data, using constraint programming. Then, we propose two mixed integer linear programmes to address this air-rail synchronisation issue, that modify flight and train initial schedules in order to offer passengers seamless connections between the two modes. The first model, well-suited for long-term planning, allows transportation schedulers to include seamless connections with other modes in their operations. The second formulation aims at rescheduling trains and flights at a tactical level to minimise the overall passenger delay across a multimodal transportation network. Both models take into account operational constraints, and are tested on the Western Europe transportation network. In a final step, we propose to rethink the scheduling process from scratch, and jointly develop an air-rail transportation network : an integrated service network design process between rail and air is proposed, considering passengers' travel option preferences and CO_2 emissions. A mixed-integer linear programming formulation of the problem is proposed. The methodology is tested on the Spanish transportation case study, using mobile phone network data as the passenger demand input.

Résumé

Le transport aérien a été développé et optimisé depuis de nombreuses années afin de proposer aux passagers un service de qualité. Cependant, la prise de conscience écologique et la congestion croissante des aéroports amènent à repenser le transport aérien. Aujourd'hui, les trains apparaissent comme une alternative pertinente aux vols de courte distance. Si les trains doivent remplacer les vols, une collaboration entre les acteurs du transport aérien et ferroviaire sera nécessaire afin de maintenir un niveau de service élevé pour les passagers. À travers cette thèse, des mécanismes de coordination entre le transport aérien et ferroviaire sont proposés afin d'améliorer la qualité des transferts entre les deux modes. La demande passager entre les avions et les trains étant aujourd'hui peu connue, et non accessible de manière publique, une première étape de simulation, basée sur de la programmation par contraintes, est tout d'abord proposée. Ensuite, deux modèles de synchronisation des horaires, utilisant les plannings existants, sont développés afin de proposer aux passagers des connexions train-vol plus fluides. Le premier, adapté pour une utilisation à l'échelle stratégique (plusieurs semaines avant les opérations), permet aux compagnies aériennes et ferroviaires de proposer aux passagers des connexions plus confortables, en termes de temps de transfert, entre les deux modes. Le second modèle a pour objectif de re-planifier les vols et les trains en temps réel, afin d'attendre les passagers impactés par un retard, minimisant ainsi le risque de connexion manquées. Les deux problèmes sont modélisés sous forme de problèmes linéaires mixtes en nombres entiers, et testés à l'échelle d'un réseau de transport européen. Enfin, nous proposons dans une dernière partie, un modèle de synchronisation plus en amont des deux précédentes méthodes : une estimation jointe des fréquences journalières des vols et des trains. Le point de vue passager est adopté, avec pour objectif principal de minimiser le temps de trajet porte-à-porte, incluant des trajets multimodaux. Le coût CO_2 du réseau de transport est également pris en compte. La méthodologie est testée sur un cas d'étude réel : le réseau de transport espagnol. Une demande réaliste est obtenue à partir de l'exploitation de données des opérateurs de téléphonie mobile.

Remerciements

Bien que ce manuscrit soit le fruit de trois ans de travail personnel, il n'aurait certainement pas abouti sans l'aide et le soutien d'un grand nombre de personnes.

Tout d'abord, merci à mon directeur de thèse, Daniel Delahaye, de m'avoir fait confiance sur un simple coup de téléphone, d'avoir partagé ses connaissances du monde de l'aviation, de m'avoir permis de voyager, et donné toutes les ressources nécessaires pour mener à bien ce travail de thèse.

Aude, merci pour ta présence tout au long de ces années, malgré la distance, le décalage horaire et ton emploi du temps chargé. Tu as su me faire sortir de ma zone de confort, me faire voir les choses sous un angle différent, et appréhender la vie avec pragmatisme. Merci pour tous les conseils que tu m'as donnés, à la fois dans le travail, mais également au-delà, qui me serviront pour le futur. Plus que ma co-directrice de thèse, tu fus un mentor pour moi.

Marcel, merci d'avoir rejoint l'aventure en cours de route. Je te suis extrêmement reconnaissante pour ta disponibilité et tout le temps que tu m'as accordé ces deux dernières années, afin de répondre à mes interrogations, qu'elles fussent d'ordre technique ou logistique. Merci également pour les très nombreuses relectures chirurgicales de mes travaux, qui ont grandement contribué à améliorer la qualité de ce manuscrit, mais certainement aussi celle de tous mes futurs écrits.

Je tenais également à remercier tous les membres de mon jury : Andrew Cook, Sophie Demassey, Denis Montaut et Cyrille Briand, pour leur relecture attentive de mes travaux, leurs questions pertinentes, ainsi que leur bienveillance, qui me permettront de garder un heureux souvenir de cette journée de soutenance.

Merci à toutes les personnes du laboratoire OPTIM et en particulier l'équipe MORO : Sonia, Catherine, Andrija, Mohand, Marcel, notre famille d'accueil pour nous, jeunes doctorants en optimisation.

Un grand merci à tous les doctorants avec qui j'ai partagé ces 3 années : Andréas, Julien, Jean-Claude, Adrien, Pierre, Céline, Alexis Brun et Alexis Brégeon, Bastien, Rémi C. et Rémi P., Thomas, Emmanuel, Denis, Eliot, Antoine, et Zahraa. Une spéciale dédicace à Geoffrey, mon « faux jumeau » de thèse. Merci pour ces fous rires des vendredis après-midi quand la fatigue se faisait sentir, les aventures de voyage à Séville, Zurich, Rennes ou Lyon ; mais également pour ton aide technique et morale, au cours de ces trois ans, qui

auraient pu être bien différents si tu n'avais pas été là.

Gracias a todo el equipo de NOMMON por haberme recibido con cariño en Madrid durante esos dos meses. Su disponibilidad, conocimientos y amabilidad me han permitido trabajar con tranquilidad y aprender muchas cosas sobre el mundo de los datos.

Merci à tous mes amis d'avoir été là le 18 avril, que ce soit en désormais en « présentiel » ou en « distanciel », parfois à l'autre bout de la Terre. J'ai mesuré la chance que j'avais d'être entourée de personnes aussi formidables.

Finalement, mes derniers remerciements s'adressent aux personnes qui ont le plus contribué à l'aboutissement de ce projet. Quentin, je ne saurais lister toutes les choses que tu as faites pour moi ces dernières années pour me permettre de travailler sereinement. Merci d'avoir été là quand ça n'allait pas, d'avoir passé des heures à me remonter le moral et me rassurer, et d'avoir toujours accepté les contraintes que cela a pu engendrer.

Enfin, je tiens à remercier ma famille. Mon oncle et ma tante pour avoir pris soin de moi toutes ces années à Toulouse, et surtout mes parents, mon frère et mes grands-parents, pour leur soutien et leur aide depuis toutes ces années. Tout cela n'aurait bien sûr jamais été possible sans vous, merci pour tout.

Table des matières

Acronymes

ADS-B Automatic Dependent Surveillance-Broadcast. 27
ARSND Air-Rail Service Network Design. 105
ARSTS Air-Rail Strategic Timetable Synchronisation. 63, 64, 68, 70
ATFM Air Traffic Flow Management. 91

CDG Paris-Charles de Gaulle airport. 51
CP Constraint Programming. 25, 48

FRA Frankfurt airport. 51

GDP Gross Domestic Product. 42
GTFS General Transit Feed Specification. 29, 31, 36, 41, 54

HSR High-Speed Rail. 29, 30, 42, 56

ILP Integer Linear Programming. 45

LPP Line Planning Problem. 8, 9, 20

MACT Maximum Acceptable Connection Time. 56, 57
MAD Madrid-Barajas airport. 51
MCT Minimum Connection Time. 56, 57, 91
MILP Mixed-Integer Linear Programming. 20, 86
MTT Minimum Turnaround Time. 31, 33

OD Origin-Destination. 9, 18, 19, 38, 41, 45

RPP Route Planning Problem. 8, 9

SESAR Single European Sky ATM Research. 3

TTP Timetable generation Problem. 8, 9

Table des figures

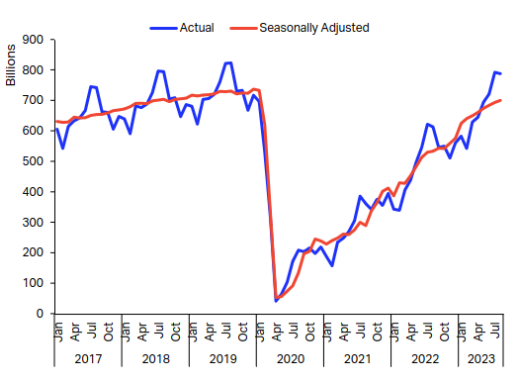
Liste des tableaux

Introduction

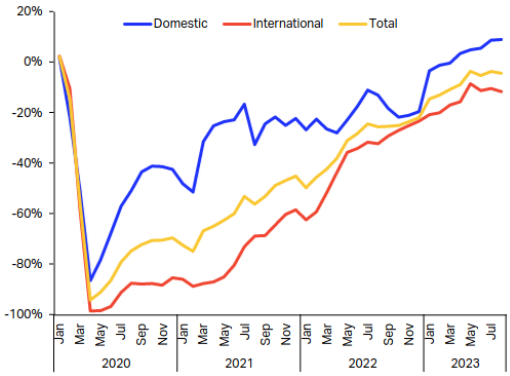
This introduction provides the context in which the thesis is developed. First, the motivation to develop coordination mechanisms between air and rail transportation modes is presented. Second, the emerging cooperation initiatives are exposed. Finally, the last subsection presents an overview of the thesis.

Motivation

The decades of steady growth in air traffic demand was interrupted by the COVID-19 outbreak. Nevertheless, recent measures show that the air passenger volume is now returning to its pre-pandemic level. One of the most widely used metrics for assessing the utilisation of a transportation mode is the Revenue-Passenger Kilometre (RPK). This metric accounts for the total number of kilometres travelled by passengers paying for transportation (*i.e.*, excluding infants not using a seat, airline agents, passengers travelling freely, etc.). Figures 1a and 1b present the total air RPK per month from January 2017 to July 2023, and its variation compared to 2019, from 2020 to 2023. In Figure 1b, a distinction is made between domestic (both origin and destination are in the country of the air carrier business headquarter) and international (origin or destination is out of the air carrier business state) markets. Air RPK volume in August 2023 reached 95% of the value of 2019. The share of domestic market overtook that of 2019. Pre-pandemic air transportation concerns are therefore coming back to the point, and the air transportation system is facing again capacity issues, leading to congestion and delays. Simultaneously, the growth of environmental awareness underscores the need to reconsider air travel. The transportation sector is responsible for 23% of the energy-related CO_2 emission (Figure 2). According to the International Energy Agency (IEA), in 2022 aviation accounted for 2% of the global CO_2 emission worldwide, with a release of 800Mt of CO_2, compared with less than 100Mt for the rail transportation system (International Energy Agency (IEA), 2023). The reduction of the environmental impact of the air transportation sector is therefore one of the main objectives of many countries in the coming years. Particularly, the European Commission set the objective that trips under 500 km should be carbon neutral in 2050 (European Commission, Directorate-General for Mobility and Transport, 2020). To facilitate the decarbonisation of the air transportation system, various policies have already been enacted.



(a) Revenue-Passenger Kilometre per month.



(b) Domestic and international RPK variation compared with 2019.

Figure 1 – Air passenger traffic analysis International Air Transport Association (IATA) (2023a).

For instance, in 2021, the French government banned short-haul flights if they could be replaced by a rail alternative in less than 2h30. During the summer 2023, the Dutch government decided to limit the annual flight volume in 2024 at Amsterdam Schiphol airport to 440,000 movements. This corresponds to a reduction of 12% compared to 2023, in order to decrease noise inconvenience and greenhouse emissions. The multiplication of such measures will force air transportation stakeholders to revise and adapt their operations. In such a context, rail transportation appears as a relevant alternative and collaborative transportation mode.

A review of air-rail collaboration initiave

Historically, air and rail transportation providers were competitors on the middle-distance market. Many studies focus on the potential competition between the two modes on several routes, and both airlines and railway operators used to consider the other mode in their operations to maintain attractiveness and their competitiveness. For instance, Behrens and Pels (2012) study the competition between high-speed rail and flights on the Paris-London market. Park and Ha (2006) analyse the impact of the development of highways and high-speed train on the air transport demand in Korea. Monmousseau *et al.* (2019) compare air and rail door-to-door travel times from Paris to London. Another important factor that limited the collaboration between air and ground transportation modes is airport locations. Generally, airports are located outside of the cities, mainly to mitigate noise disturbances and as a safety measure. They are consequently mainly accessed via private cars (Vespermann and Wald (2011)), reducing the potential for collaboration

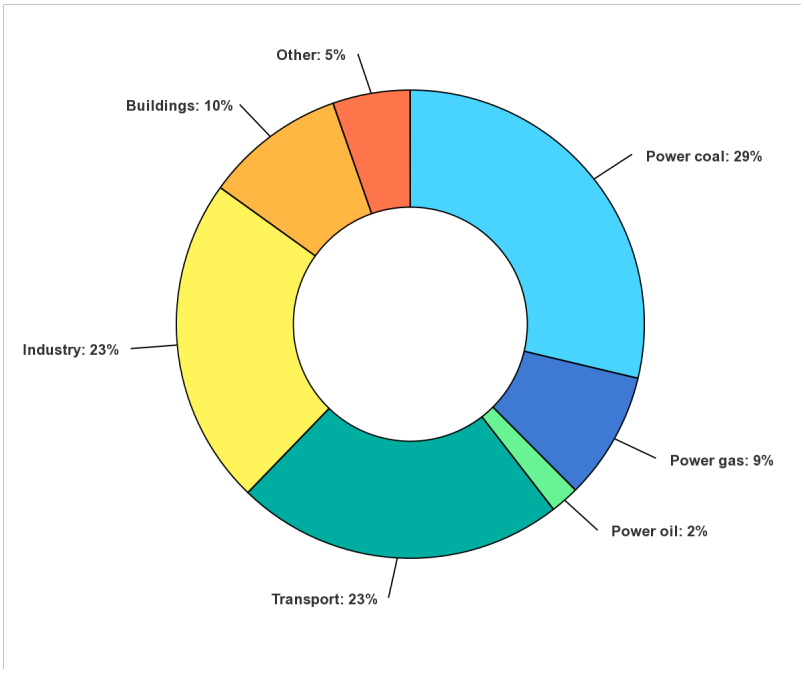


Figure 2 – Global energy-related CO_2 emissions by sector International Energy Agency (IEA) (2020).

with public transportation modes. These factors contributed to limit the development of cooperation between these two actors of long-distance transportation.

Nevertheless, the current perspective is shifting towards greater cooperation and collaboration. Indeed, Givoni and Banister (2006) are among the first authors to study collaboration instead of competition between air and rail, suggesting to replace some *feeder flights* by train to relieve airport congestion. Feeder flights, in this context, refer to short-distance flights strategically scheduled by airlines to route passengers through their hubs to connect with long-distance flights, when direct connections are not available. The multiplication of research initiatives aiming at strengthening cooperation between air and ground transportation modes underscores the evident potential for enhancing multimodal operations. In 2004, the Single European Sky ATM Research (SESAR) programme was initiated in Europe with the objective of modernising the management of the air traffic system across the continent (European Commission (2023b)). Through this program, several research projects have been conducted to develop coordination mechanisms between air and ground transportation modes. For instance, MODUS (SESAR Joint Undertaking, 2020) aimed to measure the performance of the transportation system considering the entire passenger door-to-door journey. IMHOTEP (Mota *et al.*, 2020) proposed to extend airport operations management, including collaboration with ground transportation modes. More recently, the

MultiModX project (SESAR Joint Undertaking, 2023a) was launched with the objective of designing air-rail operations considering passengers' perspective. In particular, this PhD dissertation is funded by the TRANSIT project (Bueno *et al.*, 2022). The main purpose of this European project is to develop coordination mechanisms between air and ground transportation modes both at the strategic (long-term planning) and tactical (real time) levels, to improve the passenger door-to-door journey. Finally, in October 2023, the European Commission initiated a call for proposals entitled "Integrated air and rail network backbone for a sustainable and energy-efficient multimodal transport system". This call is jointly initiated by SESAR and Europe's Rail Europe's Rail (2023), each recognised for their expertise in enhancing air and rail operations in Europe. This call underscores the readiness of both entities to collaborate closely.

In parallel to research projects, various collaboration initiatives are currently ongoing :

— Rail&Fly : An agreement between an airline (Lufthansa) and a railway operator (Deutsche Bahn) that allows air travellers to benefit from the rail network to reach or leave Frankfurt airport. More than 5600 train stations are accessible within this collaboration. However, passengers are not covered in case of missed connection due to late arrivals at the airport ;

— Lufthansa Express Rail : Trains and flights are coordinated, only one single ticket is necessary, and a guarantee of correspondence is provided. Hence, if a delay occurs, passengers are automatically registered for the next train or flight ;

— AIRail : Right after arriving at the long-distance train station at Frankfurt Airport, Lufthansa Express Rail customers can quickly and conveniently drop off their baggage and also check in for their flight in the nearby AiRail Terminal ;

— Train+Air : Integrated service ticketing between Parisian airports (Roissy and Orly) and 19 train stations in France. Passengers can purchase their tickets on the SNCF website, in travel agencies or airline partners, and use it along their whole trip as a unique ticket. Within this system, multimodal trips are proposed if the connection times between the train and the flights are sufficiently long to allow passenger transfer. These trips are proposed using the flight and train schedules defined by each supplier, but no additional schedule coordination is implemented ;

— City Airport Train in Vienna : Passengers can check-in their luggage in the city centre of Vienna. This service allows passengers to transfer easily between the city and the airport without carrying their luggage, encouraging them to use public transport instead of taxis or private cars.

These soft measures pave the way towards seamless multimodal journey ; however today, except for Lufthansa Express Rail, no schedule synchronisation is implemented. If airlines have to replace feeder flights by trains, coordination is required to maintain attractiveness and to guarantee a high-level of service to passengers. Indeed, using a combination of

transportation modes to reach one's destination is feasible only when schedules allow it. In particular, the connection time between two segments of the journey is critical for passengers. Air passengers are today protected by the European regulation 261/2004 (European Commission, 2004), which guarantee passengers compensation in case of flight delays and cancellation. However, no guarantee exists today regarding multimodal trips.

Thesis overview

In this dissertation, we consider the setting in which a passenger wants to travel from an origin to a destination, using either rail, air or a combination of both modes, with a unique ticket, in a limited amount of time and with a targeted arrival time window. The question we address is : would it be possible to design a transportation supply that could satisfy these expectations ? In particular, air and rail schedules must be synchronised to monitor transfer times and CO_2 emissions.

The first mechanism proposed in this dissertation, well-suited for medium-term planning, builds upon existing schedules developed by transportation providers. It involves marginal adjustments to these schedules, taking into account the constraints of the operators, with the ultimate aim of enhancing the seamless transfer of passengers. The second mechanism, inspired from the first one developed, has the objective to minimise the impact of delay on multimodal passenger journeys. Similarly, we propose to change slightly train and flight schedules, in order to wait for delayed passengers and reduce the number of missed connections. The third mechanism operates on a long-term scheduling horizon. It proposes to go a step further in air-rail planning integration, and to design jointly flight and train frequencies. Passengers' travel preferences are taken into account as well as the environmental cost of the transportation network. All synchronisation mechanisms proposed are based on the existing infrastructure, and no additional infrastructure cost is required.

The dissertation is organised as follows. The scheduling process employed by transportation service providers for flights and trains is initially discussed in Chapter 1. Achieving synchronisation between these two transportation modes necessitates the acquisition of data from both sides to account for operational constraints. Furthermore, since our primary goal is to enhance the passenger experience, it is imperative to estimate air-rail passenger demand, which will serve as a critical input for the synchronisation algorithms we shall introduce. The collection of data and the estimation of multimodal demand are addressed in Chapter 2. In Chapter 3, we present the two first proposed synchronisation mechanisms, which entails making minimal adjustments to existing schedules. This approach allows one to improve transfer convenience for passengers. Finally, Chapter 4 outlines the process of jointly planning the frequencies of both flights and trains, a further step in air-rail collaboration.

The publications stemmed from this dissertation and the data set generated are listed in Appendix C.

Chapitre 1

A review of transport scheduling for timetable optimisation

This chapter reviews the different steps followed by transportation operators to design their schedules. The hierarchical scheduling process is first presented with an emphasis on two particular steps : the line planning problem and the timetable generation problem. Then, schedule coordination mechanisms developed within and across transportation modes, to improve passengers transfers, are discussed.

1.1 Design of transportation schedules

This section first presents the general process of transportation network construction, followed by a detailed description of the route planning problem and the timetable generation process.

1.1.1 The overall design process

The objective of transportation operators is to provide passengers with an efficient and attractive service in order to maximise their profits. The offer directly corresponds to the schedule of flights and trains that are proposed by transportation planners. In order to align with passengers' expectations and provide an efficient transportation system, operators follow a several-year process to build their final timetables. The design of a feasible and optimal schedule is a complex task due to the high number of parameters to take into account. For instance, as depicted in Figure 1.1, Grosche (2009) presented an overview of all the information and constraints that airlines should consider while designing their schedules. Passenger demand, operational cost but also airport constraints, competition between airlines or with other modes, and working hour regulation, for instance, must be considered by operators to design the most profitable network. For that reason, transporta-

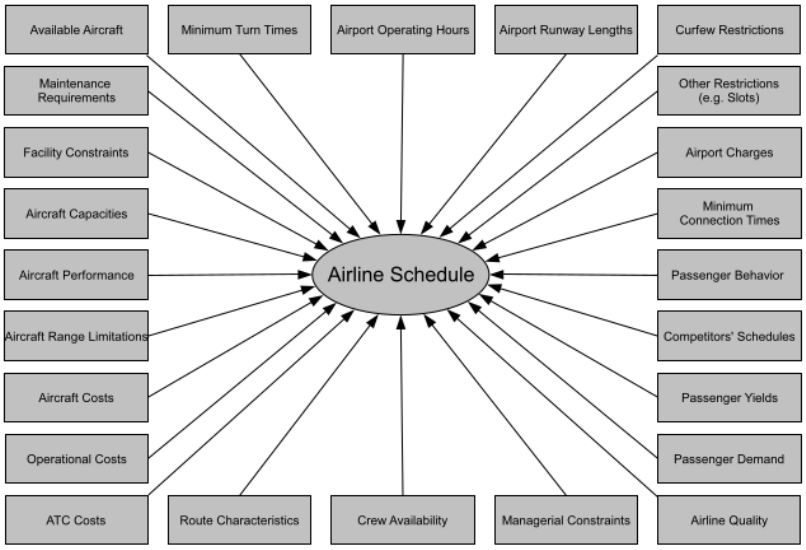


Figure 1.1 – Airline constraints to be taken into account for the construction of an optimal schedule (Grosche, 2009).

tion suppliers generally construct their timetables following an iterative procedure, referred to as the hierarchical process (Ghoseiri *et al.* (2004); Belobaba *et al.* (2009)). The main steps of such a process are presented in Figure 1.2. Long before the day of operations, transportation planners decide to build transportation infrastructures : train stations, airports, railway tracks, etc. These decisions are generally made by local or national authorities, considering many factors such as the potential market, accessibility for passengers, noise, weather or environmental impact (see Kazda and Caves (2007)). The facility location problem will not be detailed here, as it is not at the core of the dissertation, but some authors such as Phang (2003); Keeney (1973); Banister and Berechman (2003) deal with that topic. In a second step, once the infrastructure is established, transportation operators must decide on which route to operate and with what frequency. This step is referred in the literature to as the Route Planning Problem (RPP) for airline, or the Line Planning Problem (LPP) for railway operators. These decisions depend on several factors such as the potential market but also competition with other transportation suppliers. They are generally made a year or several months in advance (*strategic* planning horizon). Then, suppliers set the exact departure time and arrival time of each trip. This step corresponds to the Timetable generation Problem (TTP) and occurs months in advance. The TTP is decisive for transportation suppliers as the scheduled time of a trip directly impact the expected revenue. Finally, vehicles (aircraft or train-vehicle) and crews (pilots, train drivers, etc.) are assigned to each scheduled leg to ensure that each journey can be operated. These last steps are planned several weeks before the day of operations and are submitted to

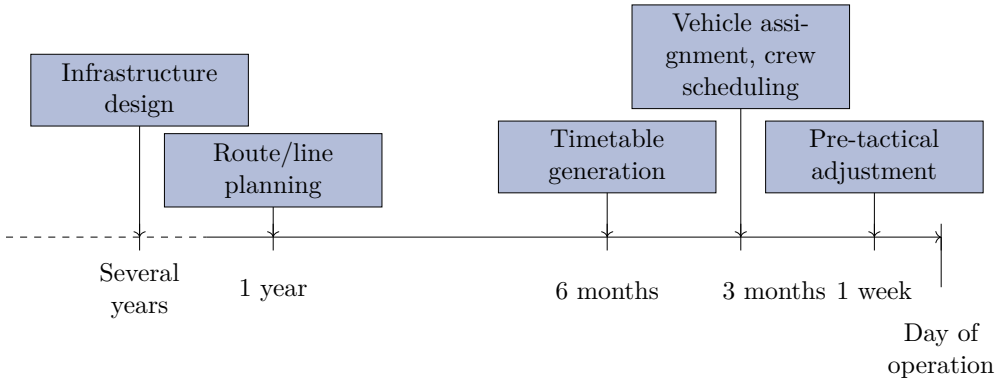


Figure 1.2 – The overall scheduling process.

pre-tactical (a few days before operations) or *tactical* (the day of operations) adjustments in case of problems. The complexity of the overall scheduling process is likely to lead to computational issues when solving real case problems. Consequently, each stage of the timetabling process is solved individually, using the result of the previous step as an input. Unfortunately, this is likely to lead to sub-optimal solutions, and some studies propose to integrate several steps at once to obtain better results (Grosche (2009); Schiewe (2020)).

In this thesis, we aim at developing synchronisation mechanisms between flights and trains in terms of timetable. We thus focus on two specific stages of the timetabling process, the RPP/LPP and the TTP. A description of these two sub-problems is presented in the next two subsections.

1.1.2 Line and frequency planning

The route or line planning problem consists in selecting at what frequency transportation suppliers will operate on each route. In other words, it amounts to deciding the daily number of flight or train journeys planned between two airports or train stations. Based on the long-term forecast of passenger demand, the decisions are made to maximise suppliers' expected profit. Bussieck *et al.* (1997) proposed a basic passenger-centric approach to reduce passenger travel time by maximising the number of *direct* travellers in a rail transportation network, *i.e.* travellers that can reach their destination station from their origin station without intermediate transfer. Their formulation of the LPP goes as follows. A transportation network is given under the form of a directed graph $G = \{\mathcal{V}, \mathcal{E}\}$, whose vertex set, $\mathcal{V}$, represents the set of stations, and the edge set, $\mathcal{E}$, gives the links between the stations. For railway operators, a *line* corresponds to a simple *path* in the network, from an Origin station $O \in \mathcal{V}$ to a Destination station, $D \in \mathcal{V}$. In other words, a line is a set of edges of the form $\{(O, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_k), (v_k, D)\}$, for some integer k. Let $\mathcal{L}$ denote the set of lines, and let $\mathcal{T}$ be the set of Origin-Destination (OD) pairs for which

the travel demand is not null. Let p_t be a shortest path for the OD pair $t = (O, D) \in \mathcal{T}$ in the graph G. For each edge $e \in \mathcal{E}$, we define $\mathcal{T}_e = \{t \in \mathcal{T} | e \in p_t\}$, the set of OD pairs t having edge e in is associated shortest path p_t. Similarly, for each line $l \in \mathcal{L}$, we define $\mathcal{T}_l = \{t \in \mathcal{T} | p_t \subseteq l\}$, the set of OD pairs t whose associated shortest path p_t is included in line l, and conversely, and we also define $\mathcal{E}_l$, the set of edges included in line l ; *i.e.*, $\mathcal{E}_l = \{e \in \mathcal{E} | e \in l\}$. For each OD pair $t \in \mathcal{T}$, we define $\mathcal{L}_t = \{l \in \mathcal{L} | p_t \subseteq l\}$ the set of lines including the associated shortest path p_t, and we are given n_t, the volume of passengers who plan to travel from station O to station D. Let C denote the maximal train capacity, *i.e.*, the maximum number of passengers who can be carried in a single train (C is a given input data).

This formulation uses, for each $t \in \mathcal{T}$, and each $l \in \mathcal{L}_t$, an integer decision variable, noted d_{tl}, that counts the number of travellers one assigns to the OD pair t on line l. It also uses an integer decision variable, noted f_l, that sets the *frequency* of line l, $l \in \mathcal{L}$. The objective of this formulation of the LPP is to maximise the number of direct travellers, and is formulated as follows :

$$\max_{d,f} \sum_{l \in \mathcal{L}} \sum_{t \in \mathcal{T}_l} d_{tl} \tag{1.1}$$

subject to :

$$\sum_{l \in \mathcal{L}_t} d_{tl} \leq n_t \qquad t \in \mathcal{T} \tag{1.1a}$$

$$\sum_{t \in \mathcal{T}_e} d_{tl} \leq C.f_l \qquad e \in \mathcal{E}_l, l \in \mathcal{L}, \tag{1.1b}$$

$$d_{tl}, f_l \in Z_+ \qquad t \in \mathcal{T}, l \in \mathcal{L}, \tag{1.1c}$$

where d is a matrix whose (t, l) component is d_{tl}, and f is a vector whose l^{th} component is f_l. Equation (1.1) counts the total number of passengers who are able to travel without any transfer within the transportation network. Constraints (1.1a) guarantee that the number of direct travellers is lower than the total number of travellers on each OD pair, and constraints (1.1b) ensure that the number of travellers do not overtake the maximum capacity constraints. Claessens *et al.* (1998) suggested a model and an algorithm to select lines for the Dutch Railways that also takes into account the operational cost. Similarly, Goossens *et al.* (2006) proposed a solution to the line planning problem that minimises the operational cost. Regarding public transportation, Ceder and Wilson (1986) proposed a methodology to design a bus network from scratch, van Nes *et al.* (1988) developed an algorithm to select routes to be included in the public transport network, and to assign frequencies to each trip.

Regarding airline operators, the route planning problem is linked to the global structure of their network. Especially, airlines have to decide the structure of the network on which they want to operate. Lederer and Nambimadom (1998) describe several possible network

structures in use, among which are the direct network, the tour network, the hub-and-spoke network, and the subtour network, illustrated respectively in Figure 1.3 (a), (b), (c) and (d). Depending on the particular airline strategy, one network can be more profitable than

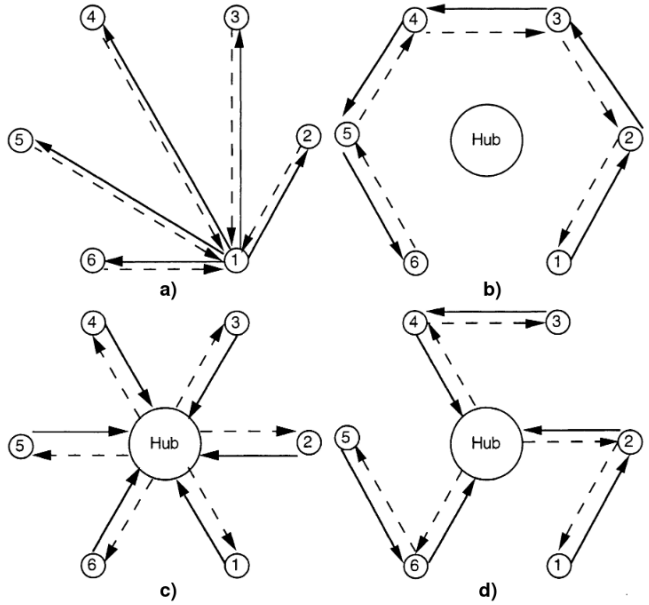


Figure 1.3 – Major network structures : direct network (a), tour network (b), hub-and-spoke network (c), and subtour network (d) (Lederer and Nambimadom, 1998).

another. Also, as explained by Lederer and Nambimadom (1998), the distance between cities, the number of cities served, and the demand rate have a direct impact on the choice of a particular network structure. The objective of airlines is therefore to determine the flight frequency on each OD pair that minimises their operational cost. Generally, the network design problem consists in routing a set of commodities, $\mathcal{K}$, at the lowest cost for transportation service providers. In case of long-distance transportation, either freight or passenger, a commodity correspond to a group of goods travelling between an OD pair. For instance, in the case of passenger transportation, a first commodity may correspond to passengers travelling from Marseille airport to London airport, another one is representing passengers travelling from Marseille airport to Toulouse airport. For each commodity $k \in \mathcal{K}$, o_k and d_k denote the origin and the destination of commodity k, respectively, and R_k, the associated traveller volume. Let $G = \{\mathcal{V}, \mathcal{A}\}$ denote the air transportation network, with the vertex index set, $\mathcal{V}$, representing the set of airports, and the arc index set, $\mathcal{A}$, the set of flight routes. For each route $(i, j) \in \mathcal{A}$, the maximum flight frequency and the fixed route design cost are denoted by K_{ij} and $F_{i,j}$, respectively. For each route $(i, j) \in \mathcal{A}$, c_{ij}^k corresponds to the unit cost of routing commodity k on (i, j). For each commodity $k \in \mathcal{K}$, and each route $(i, j) \in \mathcal{A}$, we define f_{ij}^k the continuous decision variable, indicating the flow

of k on route (i, j). For each route $(i, j) \in \mathcal{A}$, y_{ij} denotes the binary decision variable that indicates whether route (i, j) is selected or not. The network design problem formulation, inspired from Magnanti and Wong (1984), is the following :

$$\min_{f,y} \sum_{k\in\mathcal{K}} \sum_{(i,j)\in\mathcal{A}} c_{ij}^k f_{ij}^k + \sum_{(i,j)\in\mathcal{A}} F_{ij} y_{ij}, \tag{1.2}$$

subject to :

$$\sum_{j\in\mathcal{V}} f_{o_k j}^k - \sum_{l\in\mathcal{V}} f_{l o_k}^k = R_k \qquad k \in \mathcal{K}, \tag{1.2a}$$

$$\sum_{j\in\mathcal{V}} f_{i d_k}^k - \sum_{l\in\mathcal{V}} f_{l d_k}^k = -R_k \qquad k \in \mathcal{K}, \tag{1.2b}$$

$$\sum_{j\in\mathcal{V}} f_{i d_k}^k - \sum_{l\in\mathcal{V}} f_{l d_k}^k = 0 \qquad i \in \mathcal{V}\backslash\{o_k, d_k\}, k \in \mathcal{K}, \tag{1.2c}$$

$$\sum_{k\in\mathcal{K}} f_{ij}^k \leq K_{ij} y_{ij} \qquad (i, j) \in \mathcal{A} \tag{1.2d}$$

$$f_{ij}^k \geq 0 \qquad (i, j) \in \mathcal{A}, k \in \mathcal{K} \tag{1.2e}$$

$$y_{ij} \in \{0, 1\} \qquad (i, j) \in \mathcal{A} \tag{1.2f}$$

The first term of Equation (1.2) counts the cost of routing each commodity on selected arcs, while the second term counts the fixed cost of opening the routes. Constraints (1.2a)-(1.2c) are flow conservation constraints, and constraints (1.2d) ensure that the flow on route (i, j) does not exceed the maximum capacity. Nowadays, the hub-and-spoke network is the most widely used structure since it allows airlines to operate an higher frequency between city-pairs with a lower number of aircraft compared to a direct network (Belobaba *et al.*, 2009). However, the hub-and-spoke structure has a cost for passengers who have to connect at the hub to reach their final destination. Consequently, this configuration, is often less comfortable and efficient for passengers. Furthermore, an additional problem must be addressed by airlines which consists in deciding where to implement their hub, referred to as the Hub Location Problem (see Alumur and Kara (2008); Campbell (1994)).

According to Belobaba *et al.* (2009), airlines establish route profitability models to support their planning decisions. These models contain accurate information on airline expenditures for each route. Based on these estimations, airlines then decide which route to integrate in their network. Finally, airlines establish the frequency planning. Frequency is intrinsically related to competitiveness on a market. Indeed, as exposed by Simpson (1969), a higher frequency will increase the market share of an airline. Simpson (1969) proposed several models for airlines to construct their frequency planning in order to minimise their operational cost. Ghobrial *et al.* (1992) developed a heuristic to determine flight frequencies and aircraft routing for small-size airlines. Teodorović and Krčmar-Nožić (1989) developed a method to determine flight frequencies on routes subject to competition. After the selection

of routes and frequencies, transportation suppliers must fix the departure and arrival times of each trip. This step corresponds to the timetable planning development, presented in the next section.

1.1.3 Timetable development

The timetable development step corresponds to the selection of the departure time and arrival time of each leg operated by transportation suppliers. Generally, for all transportation suppliers, the timetable is repeated daily or weekly. Hence, the timetabling problem needs to be solved only for one period of time. Furthermore, an operator can consider several objectives when building its timetable, such as minimising cost operation while maximising passengers satisfaction (minimise total travel time) and robustness (resilience to disruptions). Each transportation mode has its own process to determine the final timetable. The process for airlines and railway operators are presented below.

1.1.3.1 Flight timetabling process

Historically, computational limitations did not allow airlines to consider all the operational constraints when building their timetables. The schedule design was manually handled based on an iterative process : *schedule construction* and *schedule evaluation* (Etschmaier and Mathaisel, 1985). A feasible schedule that satisfies operational constraints was established (schedule construction step), then a set of experts evaluated this schedule and suggested changes to improve it. The process was repeated until no more change was required. Heuristics were also employed to design the schedule and to speed up the process (Gopalan and Talluri, 1998). With the progress in operations research algorithms and computational power, more and more airlines use optimisation models to solve the scheduling problem. However, these methods do not build schedules from scratch. As explained by Belobaba *et al.* (2009), the construction of an entirely new timetable has some drawbacks :

— it requires recent and accurate data ;

— it is computationally costly to handle all the operational constraints ;

— it can result in hard changes compared to the original timetable, which are not desired by airlines looking for regularity valued by passengers.

Hence, most of the optimisation procedures only allow limited changes on a given schedule. They are generally coupled with a resolution of the Fleet Assignment Problem (FAP) that corresponds to assigning a fleet type to each flight. A mathematical formulation of the problem, inspired from Hane *et al.* (1995) and Belobaba *et al.* (2009), is proposed as follows.

First, here is the given input data. Let $\mathcal{L}$ be the set of scheduled flights, $\mathcal{F}$ be the set of aircraft fleet types, and, for each fleet type $f \in \mathcal{F}$, let n_f be the available number of aircraft of that fleet type. The model is based on a time-space network representation.

More precisely, a node in the graph corresponds to an event, either departure or arrival, at a given airport, as illustrated in Figure 1.4. Arcs correspond either to flights, or ground

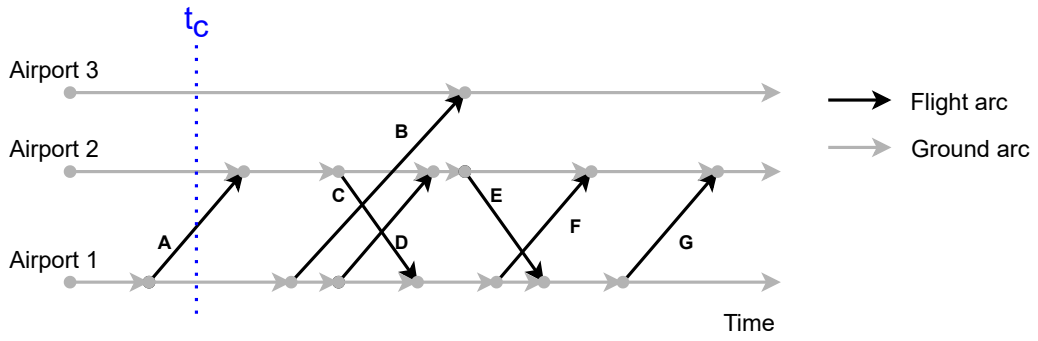


Figure 1.4 – Example of the time expanded network, with nodes corresponding to the arrival or departure of a flight, and edges correspond to flight arcs (black) or ground arcs (gray). The count time line is represented by the blue dashed line.

times at the airport between two consecutive flights. For each fleet type $f \in \mathcal{F}$, such a time-space network is defined. Indeed, depending on the fleet type, the flight time and the turnaround time between two flights (*i.e.,* time between two consecutive flights operated by a same aircraft) can vary. Therefore, a given scheduled flight $l \in \mathcal{L}$ can have a slightly different departure time or arrival time from one fleet-type network to another. Let $\mathcal{N}^f$ denote the set of nodes and G^f the set of ground arcs of fleet type f. We are given a time of the day, called the *count time*, and that we note t_c, a user-defined fixed parameter. Let $C^f_{\mathcal{L}}$ denote the set of flight arcs of $f \in \mathcal{F}$ whose departure times are before t_c and the arrival times are after t_c (*e.g.*, flight arc A in Figure 1.4). Similarly, let C^f_G denote the set of ground arcs of fleet type f whose starting node time is before t_c and ending node time is after t_c. For each fleet type f, and each node $n \in \mathcal{N}^f$, we further define O^f_n and D^f_n, the sets of flight arcs whose origin is node n and destination is node n, respectively. Let g^+_n denote the unique ground arc terminating at n, and g^-_n denote the unique ground arc starting from n, $n \in \mathcal{N}^f$.

Based on these notations and input data, we define x^f_l, the binary decision variable indicating whether flight $l \in \mathcal{L}$ is flown by fleet $f \in \mathcal{F}$. Then, for each fleet type $f \in \mathcal{F}$, and each ground arc $g \in G^f$, we define an integer decision variable, y^f_g, that counts the number of aircraft of type f on the ground arc g.

The mathematical formulation of the optimisation problem therefore reads :

$$\min_{x,y} \sum_{l \in \mathcal{L}} \sum_{f \in \mathcal{F}} c^f_l x^f_l, \tag{1.3}$$

subject to :

$$\sum_{f\in\mathcal{F}} x_l^f = 1 \qquad l \in \mathcal{L} \tag{1.3a}$$

$$\sum_{l\in O_n^f} x_l^f + y_{g_n^+}^f - \sum_{l\in D_n^f} x_l^f - y_{g_n^-}^f = 0 \qquad n \in \mathcal{N}^f, f \in \mathcal{F}, \tag{1.3b}$$

$$\sum_{l\in C_{\mathcal{L}}^f} x_l^f + \sum_{g\in C_G^f} y_g^f \leq n_f \qquad f \in \mathcal{F}, l \in \mathcal{L} \tag{1.3c}$$

$$y_g^f \in \mathbb{N} \qquad g \in G^f, f \in \mathcal{F}, \tag{1.3d}$$

$$x_l^f \in \{0,1\} \qquad l \in \mathcal{L}, f \in \mathcal{F}, \tag{1.3e}$$

where x is the vector of optimisation whose (f,l) component is x_l^f, y is the vector of optimisation whose (f,g) component is y_g^f, and c_l^f is the cost of assigning fleet type f to flight l, $f \in \mathcal{F}$, $l \in \mathcal{L}$. Remark that one can easily provide an upper bound on each of the integer variables y_g^f. Equation (1.3) is therefore the sum of the cost of the assignment of a fleet type to each flight. Constraints (1.3a) ensure that each flight is covered by one fleet type and consequently linking the individual fleet-time-expanded graph together. Constraints (1.3b) are flow conservation constraints. Constraints (1.3c) ensure that no more aircraft than available on each fleet can be assigned.

Each fleet type has its own characteristics, such as a maximum capacity and minimum turnaround time. Hence, the schedule directly affects the fleeting decision. For instance, a specific departure time can be more attractive for passengers and assigning a long-range aircraft is likely to be more profitable for airlines than a smaller one that could induce *spill cost* (loss in revenue since some passengers are not accommodated due to a demand higher than the capacity). Also, changing the schedule modifies the available turnaround time between two consecutive flights and questions the initial fleeting decision. Levin (1969) was one of the first authors to propose to solve the fleet assignment problem allowing changes in the schedule if this could reduce the fleeting size. Rexing *et al.* (2000) noticed that small changes in the initial schedule could reduce the cost of the fleet assignment. This resulted in a reduction of two in the number of aircraft needed in a real-life instance he addressed, and in a cost saving of $50 million. Desaulniers *et al.* (1997) presented two formulations of the the Daily Aircraft Routing and Scheduling Problem (DARSP) and solved instances of these formulations using branch-and-bound techniques.

1.1.3.2 Train timetabling generation

Regarding the train timetable problem, two approaches exist in the literature : the *cyclic* scheduling and the *acyclic* scheduling. The cyclic scheduling consists in designing a timetable for a given period of time T, and repeating it for each following period. One of the main advantages of such an approach is that it is easy for passengers to remember the

schedule, as it is repeated every T period of time. This approach has been introduced by Serafini and Ukovich (1989) and is referred to as the Periodic Event Scheduling Problem (PESP).

An event i corresponds to an arrival or a departure of a train at/from the train station. Let E denote the set of events. Train scheduling requires the satisfaction of several constraints such as :

— *dwell time* at stations : train should stop a minimal amount of time at each train station for dropping and picking up passengers ;

— *turnaround time* at terminus : when a train reaches its destination station, a certain amount of time should be planned to ensure the cleaning of the train, the rotation of the crew, etc. ;

— *headway* constraint : a minimum time interval between two trains running on the same track should be guaranteed for safety ;

— *trip time* : the duration of a train trip should not exceed a certain limit.

A constraint links two events $(i, j) \in E \times E$, $i \neq j$. For instance, let i denote the event representing the arrival of a train at a given train station, and j the departure of the same train from the train station. It must be ensured that the departure of the train occurs after its arrival. Similarly, the time the train spends at a station is limited. The departure time from the station cannot be later than a certain time after the arrival time of the train. For each event $(i, j) \in E \times E, i \neq j$, let u_{ij} and l_{ij} be the upper bound and lower bound of the duration between events i and j, respectively. These bounds are given as input data. For each event $i \in E$, let t_i be the integer decision variable representing the scheduled time of event i. As proposed by Peeters (2003), a generic formulation of the optimisation problem is :

$$\min_{t} \sum_{i \in E} F(t_i) \tag{1.4}$$

subject to :

$$(t_i - t_j) \mathrm{mod}(T) \leq u_{ij} \qquad (i, j) \in E \times E, i \neq j, \tag{1.4a}$$

$$l_{ij} \leq (t_i - t_j) \mathrm{mod}(T) \qquad (i, j) \in E \times E, i \neq j \tag{1.4b}$$

where l_{ij} and u_{ij} represent respectively the minimal and maximal durations between events i and j, and F is a completely general objective function that computes the cost of timetable t. Several authors focus on this model such as Nachtigall and Voget (1996) or Odijk (1996).

The second approach, the *acyclic* scheduling, schedules trains for the total running period rather than following a periodic scheme. Indeed, cyclic timetabling implies that the schedule remains the same during peak hours and off-peak hours. A trade-off should therefore be found to limit congestion during peak hour while limiting the operational cost during low-demand periods (Caprara *et al.*, 2002). The acyclic timetabling has been

developed to address this issue. A mathematical formulation of the acyclic train timetabling problem is proposed by Caprara *et al.* (2002) and is presented here. Let $\mathcal{N}$ denote the set of train stations. The train timetable is represented by an acyclic graph $G = (V, A)$, where the

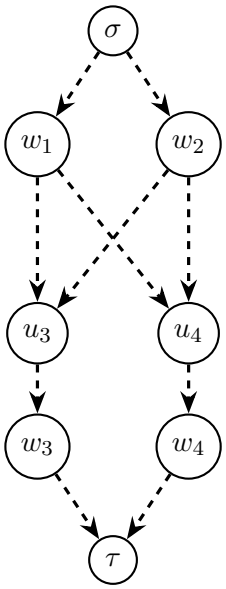


Figure 1.5 – Railway network modelling example, where σ and τ represent artificial source and sink nodes, respectively, and u_i and w_i the set of departing and arriving instants, respectively (Caprara *et al.*, 2002).

vertex set, V, corresponds to the set of possible arrival times and departure times of trains at/from train stations, and the set of arcs, A, represents either train stops at stations or train trips between two stations (see Figure 1.5). Therefore, the set of vertices is partitioned into the set of arriving instants, U, and the set of departing instants, W, an artificial source node σ, and an artificial sink node τ. The set of arriving instants U is further partitioned into subsets U^s, for each train station $s \in \mathcal{N}$, and similarly for departing-instant subset W. For each vertex $v \in V$, and each train $t \in \mathcal{T}$, let $\delta_t^+(v)$ and $\delta_t^-(v)$ denote the set of arcs in A^t respectively leaving and entering vertex v. The set of trains to be scheduled is denoted $\mathcal{T}$ and for each train $t \in \mathcal{T}$, $A^t \in A$ denotes the set of arcs that can be used by t. In order to satisfy the constraints previously introduced, we define C a subset of *pairwise incompatible* arcs. For instance, arc a_1 corresponds to the trip of train $t \in \mathcal{T}$ departing at 12 :00 from station $s \in \mathcal{N}$ and arriving at station $s + 1$ at 13 :00. Arc a_2 represents the trip of the same train t departing from station $s+1$ at 12 :50 and arriving at station $s+2$. These two arcs are incompatible since t could not depart from $s + 1$ before arriving at it. The set $\mathcal{C}$ corresponds to the set of all subsets of incompatible arcs. For each arc $a \in A$, the binary decision variable x_a indicate whether arc a is selected or not. The objective

consists in determining, in the network, the subset of arcs together with the train assigned to each selected arc that will maximise the operator profit. The optimisation problem is formulated as follows :

$$\max_{x} \sum_{t \in \mathcal{T}} \sum_{a \in A^t} p_a x_a \tag{1.5}$$

subject to :

$$\sum_{a \in \delta_t^+(\sigma)} x_a \leq 1 \qquad t \in \mathcal{T} \tag{1.5a}$$

$$\sum_{a \in \delta_t^+(v)} x_a = \sum_{a \in \delta_t^-(v)} x_a \qquad t \in \mathcal{T}, v \in V \backslash \{\sigma, \tau\} \tag{1.5b}$$

$$\sum_{a \in C} x_a \leq 1 \qquad C \in \mathcal{C} \tag{1.5c}$$

$$x_a \in \{0, 1\} \qquad a \in A \tag{1.5d}$$

where p_a is the profit associated for arc $a \in A^t$ for a train $t \in \mathcal{T}$, and x is a vector whose the a^{th} component is x_a. Constraints (1.5a) ensure that at most one arc is selected for train t to leave the source node σ. Constraints (1.5b) are flow conservation constraints. Finally, constraints (1.5c) impose that no incompatible arcs are selected. Brännlund *et al.* (1998) proposed an optimisation technique based on Lagrangian relaxation to solve the single track time-tabling problem. Wong *et al.* (2008) proposed a mixed-integer programming formulation for the timetable synchronisation problem that satisfies track constraints. Their objective is to minimise the total passenger transfer wait times by adjusting the departure and arrival times, the run time on the track and the dwell time at each station for each train.

1.2 Timetable synchronisation

As direct travels can not always be offered to passengers, some journeys require transfers between flights or trains. Indeed, the hub-and-spoke network structure generally used by airlines to reduce their operating cost involves that passengers transfer between two flights if their OD pair is not served by a direct flight. To improve the quality of such transfers and to remain attractive for passengers, transportation suppliers must synchronise their schedules. These mechanisms have mostly been developed for each individual network : airlines focus on the coordination of their flights, and railway operators concentrate on coordinating their trains. Nevertheless, multimodal timetable synchronisation models are developed, especially in the case of short distance public transport or for freight transportation. Some initiatives are now flourishing to synchronise air and rail timetables.

The next subsection presents, synchronisation mechanisms within the same transportation mode, either air or rail. Then, subsection 1.2.2 introduces intermodal synchronisation techniques.

1.2.1 Intramodal timetable synchronisation

Regarding airlines operations, hub-and-spoke networks have been developed because they allow operators to serve a large number of destination with a lower number of aircraft, compared to point-to-point network (Belobaba *et al.*, 2009). However, as mentioned above, this structure has some drawbacks for passengers who have to shift between two flights when their OD pair is not served by a direct flight. Generally, airlines operate in connecting *bank* structure at airports : a group of arriving flights is followed by a group of departing flights in order to maximise connection opportunities for passengers. The quality of connections at hub airports is therefore a factor of competitiveness for operators. This quality, referred to as *temporal connectivity*, is bounded due to airport capacities ; indeed, the number of aircraft movements per time interval is limited for safety purpose. Several studies focus on quantifying flight connectivity at airports. Burghouwt and Redondi (2013) summarise and compare different connectivity metrics used in air transportation to assess the quality of schedule coordination at hub airports. Many of these metrics consider connections involving a transfer time that ranges within predefined minimal and maximal connection times. Veldhuis (1997) defines a *connectivity unit value* that measures the quality of the indirect connection compared with a direct connection. Transfer times are penalised as being perceived three times longer than in-vehicle time. Dennis (1994) proposes a *connectivity ratio index* that compares the number feasible connections within a 90-minute range of a given schedule with a schedule of flights randomly distributed across the day. A connection is considered to be feasible if it respects a so called *Minimum Connection Time* (MCT). Burghouwt and de Wit (2005) define the *weighted indirect connection index*. They consider both the connection time quality and the detour generated by stopping at the hub airport. The connection time quality is a linear function decreasing from 1 to 0 for times between MCT and a so called *Maximum Acceptable Connection Time* (MACT). Budde *et al.* (2008) propose to measure the quality of connectivity at hub airports by taking into account the weekly frequency of flights. The authors assume that a low weekly frequency should be counterbalanced by a short connecting time. However, even if short connections are favoured by passengers to limit the total travel time, short connections are not robust in case of delay. It is a source of stress for passengers who prefer longer connection times to be more conservative. In that sense, Danesi (2006) defines an intermediate connection time that lies between *short* and *long* connections, where *short* and *long* connections are those with a connection time under and above the intermediate connection time, respectively. Theis *et al.* (2006) study passenger itinerary choice according to the connecting time between two flights. These authors assume that the utility function of the connecting time is an *n-shape* curve : short-connection time are not desired due to a feeling of rush and risk

of missed connection, long-connecting time are perceived as a loss of time for passengers. They test their hypothesis by analysing results of a survey led on approximately 5000 observations. Their results show that indeed, passengers do not attempt to minimise simply the connection time but are rather looking for a "reasonable" connection time.

Regarding railway operators, the LPP aims at designing rail lines in order to maximise the number of direct travellers. For travellers that cannot benefit from a direct connection, timetable synchronisation is required when timetables are designed. Generally, the objective of railway operators is to minimise passenger wait times at transfer stations. This mechanism is mainly developed in most of ground public transport systems (bus, subway, rail, urban rail transit, etc.). Nachtigall and Voget (1996) present a genetic algorithm to minimise passenger wait times at transfer stations in a periodic schedule. Wong *et al.* (2008) propose a method that minimises transfer times within a rail mass transit network. Jansen *et al.* (2002) propose a tabu search algorithm to minimise passenger transfer times in a public transport network. Another synchronisation approach is that of Ceder *et al.* (2001) who propose to maximise the number of simultaneous bus arrivals at stations. They formulate this problem as a Mixed-Integer Linear Programming (MILP) problem and propose a heuristic to solve it approximately in polynomial time. Similarly, Cao *et al.* (2019) develop a methodology to maximise the number of transfers at railway station, considering passenger transfer times to switch between lines. They test their methodology on the Beijing subway network, increasing by 67% the number of synchronisations compared with a basic planning (constant headway and same initial departure times at 00 :00 for all train lines). Vansteenwegen and Van Oudheusden (2006) propose to compute ideal buffer times of trains to minimise passenger wait times due to missed connections in case of delays. More precisely, the authors estimate the distribution of train actual arrival times (with potential delays), to include an additional buffer times in the train schedule so as to allow passengers to make the connection, even in case of delays.

As observed, most transportation service providers manage schedule synchronisation within their own network. There are several reasons for this. First, the lack of data sharing and of standardisation between transportation modes limits the potential for coordination. Second, there are few incentives for transportation suppliers to alter their network performance in order to synchronise with other modes (Xia and Zhang, 2017). Without a bilateral agreement, the multimodal synchronisation can be improved but at the expense of one transportation network only. Finally, transportation suppliers cannot act on the other transportation modes ; each transportation supplier manages solely its own network. All these reasons are limitations to multimodal synchronisation. However, the trend and policy support to provide passengers with smooth transfers between modes push transportation providers to initiate collaboration.

1.2.2 Intermodal synchronisation

In this dissertation, the emphasis is put on long-distance travels, which include air transportation. However, urban and freight mobility received higher attention than long-distance multimodality in the literature.

SteadieSeifi *et al.* (2014) lead a survey on multimodal freight transportation planning. As detailed in SteadieSeifi *et al.* (2014), the planning decision refers to the *service network design* : a process that gathers the frequency planning step, the capacity allocation, the equipment planning, and the routing of commodities. Andersen *et al.* (2009) develop a model to coordinate departure times of a transportation service (rail) with external service (ship) to minimise the delivery time of goods. Puettmann and Stadtler (2010) propose a methodology to synchronise the schedules of two carriers that interact through a long-haul transportation network. They aim at building a long-haul line schedule that minimises the operational cost, based on the shipment demand from the two carriers.

Regarding urban mobility, coordination between public and private transits has been extensively investigated. For instance, Tirachini *et al.* (2014) try to find the best bus frequency, size and fares that maximise passenger welfare. They build a trade-off between crowded buses and congestion at bus stations. Hence, they both take into account passenger travel time (frequency) but also try to mitigate road congestion caused by public transport. Castelli *et al.* (2004) propose a model and an algorithm to build a multimodal schedule that minimises passenger transfer times and the operational cost. While freight and urban multimodality have been widely studied in the literature, research studies addressing both air transportation system and ground transportation are only recently begun.

The first coordination mechanism with another transportation operator appears within the air transportation system through the airline alliance system. An alliance refers to a consortium of airlines that enter into a collaborative agreement. One of the most widely adopted coordination mechanisms within these alliances is the *code-sharing* system : airlines within the same alliance can offer passengers flights operated by another airline, with the same flight code. This agreement enables airlines to expand their offerings by introducing destinations not originally served, or by providing additional flight combinations, in terms of schedules. In that sense, Yan and Chen (2007) propose a model to solve the fleet routing and flight scheduling problems jointly for two airlines. Two types of collaboration are modelled : *parallel alliance*, in which one flight of each alliance can be planned at the same time for the same OD pair, and *complementary alliance*, in which a first flight is scheduled by one airline from a first station to an intermediate one, and another flight is scheduled by another airline from the intermediate station to the destination station. Results show that a combination of these two models significantly reduces the operational costs of airlines.

Regarding air-rail timetable synchronisation, at the moment, most of the studies analyse the effect of such a collaboration on passengers, or the passenger willingness to pay for such mechanisms. For instance, Román and Martín (2014) study passenger preferences in the case of air-rail integration. They define several characteristics of coordination, such as

the connecting time, travel cost, travel time, fare integration, or luggage handling. Results show that passengers value total travel times, connecting times and airport access times. Li and Sheng (2016) study passenger travel demand for air-rail integration in China. Results show that a large connection time between air and rail (above 120 minutes) is not attractive for passengers, reducing thereby the demand for air-rail service. Chiambaretto *et al.* (2013) study the passenger willingness to pay for intermodal service. They show that "short" connection times (around 90 minutes) are valued by business travellers while longer connection time (*e.g.*, 210 minutes) are preferred by leisure travellers.

Despite the challenges, synchronising solutions between ground and transportation modes have emerged. Huang *et al.* (2021) aim at solving the *last train timetabling problem* in urban rail transit network by incorporating synchronisation with other modes such as trains and flights and not only within the rail transit network. The purpose is to guarantee passengers who have a train or a flight late at night, to transfer within the urban transit network. Marzuoli *et al.* (2016) study the benefit of coordination between air and ground transportation systems in case of a major disruption. They develop an optimisation process to mitigate disruptions on passengers in case of failure of the air transportation system. They propose to accommodate stranded passengers with ground transportation modes to reach their final destination. The methodology is tested on the case of the Asiana crash, and successfully results in relocating all passengers within six hours of delay, compared to several days without considering buses. More recently, Ke *et al.* (2020) aim at synchronising trains and flights at a specific airport. The objective is to maximise the number of feasible connections between trains and flights, by changing the rail schedule. They also consider maximising the number of synchronised flights and passenger satisfaction regarding transfer times. Since these three objectives do not have the same priority, an *a priori* method is employed to solve this multi-objective optimisation problem. The above approaches propose tactical and strategic adjustments of air and rail schedules to improve passengers experience. A step further in collaboration can be reached by synchronising air and rail at the frequency planning-level. For instance, Allard and Moura (2014) introduce a model to design an hub-an-spoke structure, combining trains and flights. Okumura and Tsukai (2007) propose to determine the optimal flight frequencies and train speeds under various airport capacities and rail-line length constraints in Japan, while keeping the initial schedule train frequencies.

1.3 Conclusion

This literature review aims at giving an overview of the scheduling process that airlines and railway operators follow. This process is sequential, decomposed into several subproblems : the infrastructure design and building, the route planning, the timetable development, the vehicle scheduling, and the crew scheduling. These sub-problems are generally solved independently and without taking the other modes of transportation into

consideration. In addition, according to the literature, schedules are never designed from scratch. Flight schedules are adjusted seasonally to adapt to the demand, and train timetables are adjusted to ensure feasibility and track-constraint satisfaction. Nevertheless, the literature shows that some recent initiatives to coordinate transportation modes have emerged, with the purpose of improving the passenger door-to-door journeys. Initially developed for freight and urban transit, synchronisation models between flight and train are appearing. However, several limitations remain. First, a large part of the studies focus on the effect of potential synchronisation mechanisms on passengers (*e.g.,* several values of transfer time, luggage handling, attractive costs, etc.), but few authors actually developed synchronisation models that yield timetables that can be implemented in operational contexts. They are generally limited to consider the other modes as a constraint (Ke *et al.*, 2020), leading to unilateral changes in the schedule and potentially unfair solutions. Finally, synchronisation mechanisms are generally applied at one airport or at the national scale.

This dissertation tackles the following problem : can synchronisation models at a large scale (covering several airports/train stations or countries) improve operations for transportation suppliers, passengers and the environment ? The next chapter consists in gathering both supply and demand data as to build realistic instances as inputs for new synchronisation models.

Chapitre 2

Data collection and intermodal passenger demand simulation

In order to build their timetables, transportation operators rely on two separate data sources : supply and demand. Supply data correspond to the number of vehicles available (aircraft and trains), crews, pilots, etc. Demand data correspond to the expected number of passengers who travel on the transportation network considered.

Long-distance transportation suppliers, such as airlines or railway operators, manage only one mode of transportation, with information related to their network, exclusively. Such data is usually private and not publicly accessible since it might ensure a high-level of competitiveness. As observed in the previous chapter, the first challenge when synchronising air and rail schedules is to collect suppliers' data to have accurate knowledge about their resources. In addition, timetable generation is intrinsically related to passenger demand. Indeed, operators tend to maximise their profit when satisfying the transportation demand. This demand is estimated by suppliers using forecast models (generally relying on historical data and exogenous information) that are also private. Moreover, such predictions are usually limited to one transportation mode, and suppliers have no information on the passengers' itineraries when they exit their own network. Consequently, one of the main challenges when building an integrated timetable between air and rail is to estimate the transfer demand between these two modes.

This chapter deals with the data collection of transportation suppliers and the intermodal passenger demand simulation. First, the data sets used in the remainder of this dissertation are presented. The first section describes their collection from several sources, fusion and standardisation. Then, a data-driven intermodal passenger demand is proposed in the second section. Instances are generated using real transportation data and Constraint Programming (CP). The overall data processing scheme followed in this chapter is summarised in Figure 2.1. In the remainder of the manuscript, the term *leg* will refer either to a train journey or to a flight.

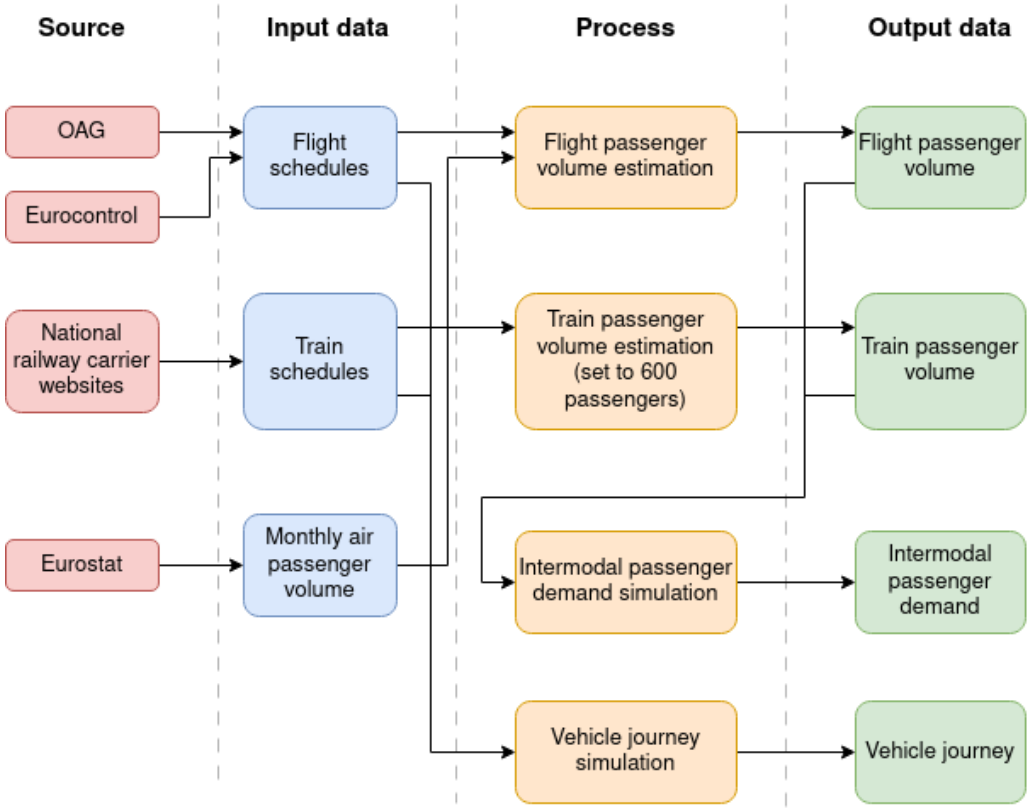


Figure 2.1 – Overall data processing scheme. The output data column presents the realistic data sets that have been simulated in Chapter 2 via the simulation processes developed in this thesis : the flight passenger volume estimation and the vehicle journey simulation, both based on heuristics, and the intermodal passenger demand simulation, obtained through a constraint-programming resolution approach applied to the optimisation model we introduced.

2.1 Data collection and standardisation

One challenge in this work was to gather accurate supply data to later propose realistic synchronisation mechanisms. In this dissertation, the first mechanism proposed relies on the existing timetables. Consequently, flight and train schedules are required input data. In this section, the collection of these inputs, from several sources, is presented. Up to date, there is no regulation regarding the sharing of transportation data and their format. Indeed, transportation suppliers are not obliged to share their data, and if they decided to do so, the format is not imposed. Therefore, as the format may differ depending on the transportation mode or even on the source, and as some essential information may be missing, we propose heuristics to realistically simulate missing data.

2.1.1 Data sources

When a passenger plans a flight trip, he or she can obtain information on flight schedules directly on airlines' websites or from an external one such as Google Flight, Kayak, etc. These platforms propose to manage passenger reservations using Global Distribution Systems. In few words, these systems are connected to several databases to provide passengers with flight information such as price, schedules or availability. The best-known data bases are Sabre (2023) and OAG (2023). In this dissertation, we aim at synchronising trains with flights considering a large number of itineraries all at once. Therefore, acquiring, in this way, flight information across various airports at national or European levels can be laborious, particularly when examining each individual origin-destination pairing. Other data sources must therefore be explored.

Daily flight schedules at large scale are usually published on airports' website or on flight tracking platform such as Flightradar24 (2023) or FlightStats (2023). These open-access tools allow one to obtain real-time information on flight status, combining several data sources (Automatic Dependent Surveillance-Broadcast (ADS-B), Multilateration, radar data, etc.) to capture the real-time position of each aircraft. However, gathering large scale information (covering several airports or countries) is a difficult task as these websites only allow information retrieval on a per-itinerary or per-airport basis. In addition, historical schedules are generally not accessible without subscription. Another accessible source of flight data is OpenSky Network (2023). This platform collects huge amounts of ADS-B data each day, storing aircraft trajectories for each recorded day. However, in our case, such data is not accurate enough. Indeed, as depicted in Figure 2.2, a flight is composed of several phases :

1. the aircraft is **at the gate**, pilots and crews prepare the flight, passengers board and the aircraft is filled with luggage and fuelled ;
2. when it is ready to go, the aircraft leave the gate (off-block time) and head towards the runway, this steps is referred to as ***taxi*** ;
3. when the runway is clear, the aircraft can **take off** (take-off time) ;

4. the **flying** phase ;
5. the aircraft **lands** at the destination airport (landing time) then proceeds (taxi) to the arriving gate, where it stops (in-block time) ;
6. the aircraft is **at the gate**, passengers can leave the aircraft and it is prepared for the next flight.

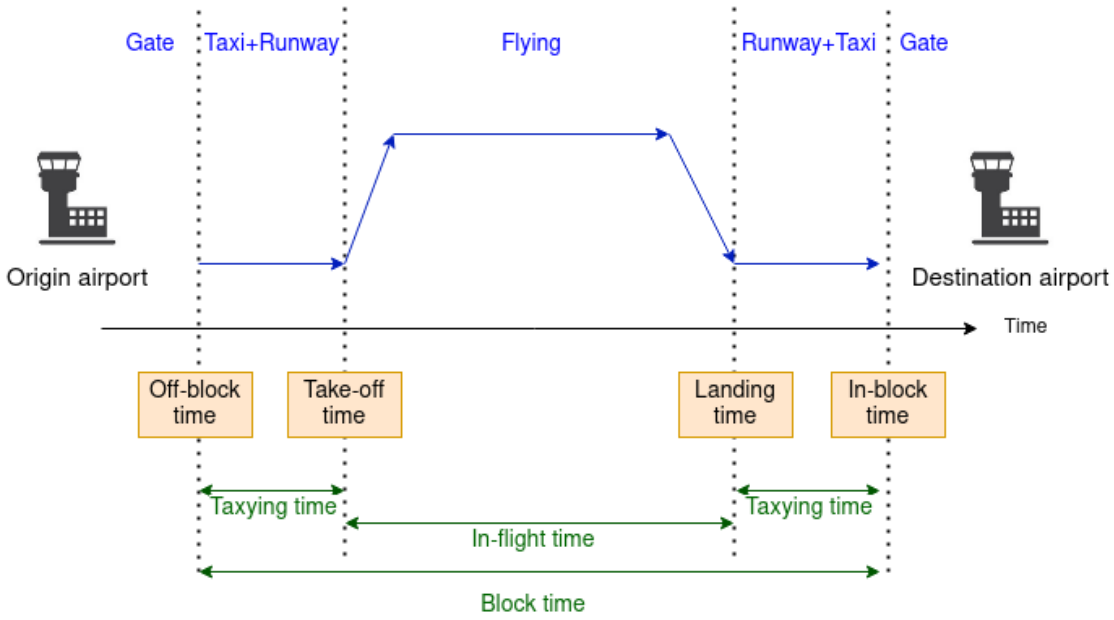


Figure 2.2 – Flight phases from origin airport to destination airport.

The departure time can either corresponds to the off-block time or to the take-off time. Similarly for the arrival time : it may correspond to the landing time or to the in-block time. To avoid confusion, in the remainder of this dissertation, departure and arrival times will refer to the off-block time and to the in-block time, respectively. Indeed, these times are the one shared by airlines with passengers when they book a flight. In addition, there might be significant differences between the off-block and departures times, or between the landing and in-block times at large airports : the taxi may last several minutes if the gate is located far from the runway or in the case of an aircraft queuing at congested hours. Therefore, it is more robust for passengers to have an idea of the time at which the aircraft will leave the gate or at what time they can disembark : they can plan their itineraries according to this information. Therefore, it is essential for us to work rather with off-block and in-block times. With Opensky data, the first and last time stamps recorded correspond to the first and last transponder emissions from the aircraft. Generally, these events occur during the taxi phase, and there is not guarantee that the first and last times recorded are the time scheduled by airlines, especially in the case of a delay. To overcome that issue, we

use another database : Eurocontrol R&D data (Eurocontrol, 2023). In Europe, Eurocontrol publishes historical flight data over Europe on a quarterly basis. Such data corresponds to flight records with either the origin and/or the destination within Europe. However, due to privacy restrictions, data is released with a two-year delay. For more recent and future schedules, data from OAG Schedule Analyzer (OAG, 2023) is used. Note however that data is only accessible through commercial subscription.

It is therefore noteworthy that currently, there is no open data repository of future flight schedules or historical ones dating back less than two years. Accessing such data proves to be resource-intensive, requiring either significant time or financial investment. An alternative approach might involve automated web scraping from individual websites ; however, this method can encounter access denial after an excessive number of requests and may be prohibited.

Regarding train schedules, most of railway operators now share their schedules under the General Transit Feed Specification (GTFS) format (MobilityData, 2023). This format provides a standardisation norm for the release of public transportation data. Historically, this system was developed in Portland, Oregon, for the release of public transportation data under an easily accessible format. Information such as stations served, stop times at the station, and stop durations are provided for the upcoming month. In Europe, national railway operators such as Société Nationale des Chemins de Fer français (SNCF) in France, Red Nacional de los Ferrocarriles Españoles (RENFE) in Spain or Deutsche Bahn (DB) in Germany now publicly share their train timetables under that format. Note, however, that historical data is not always accessible. For instance, SNCF only shares the historical data of up to 6 months, and DB only for the upcoming month. Some third-part actors now offer to store historical data. For instance, Interline, through the Transitland product Interline (2022) collects GTFS data released since 2014, and share it without fee. The International Union or Railway (UIC) developed a database gathering all rail schedules of European countries on a data base named MERITS (UIC-International Union of Railways, 2023a). However, access to this data base costs several thousands of euros per year. As this cost may be handled by private companies, it is generally not suitable for academic endeavours. To freely obtain access to the railway schedule data in Europe, one must extract the GTFS data from each operator. This could imply visiting several websites if more than one suppliers operates in a given country. In addition, railway operators generally work on a national market. The language is therefore another barrier to overcome when gathering railway data. For these reasons, we limit the scope of our study to three European countries : France, Spain and Germany.

In addition, synchronisation is relevant if there is a sufficient demand to coordinate flight and rail schedules. Consequently, we therefore focus on three hubs across each country : Paris-Charles de Gaulle (CDG), Frankfurt (FRA) and Madrid-Barajas (MAD) airports. CDG and FRA airports have a direct access to one High-Speed Rail (HSR) station directly located at the airport. Conversely, MAD airport is not equipped with a train station, passengers using train as feeder flights should therefore stop at Madrid-Chamartin or

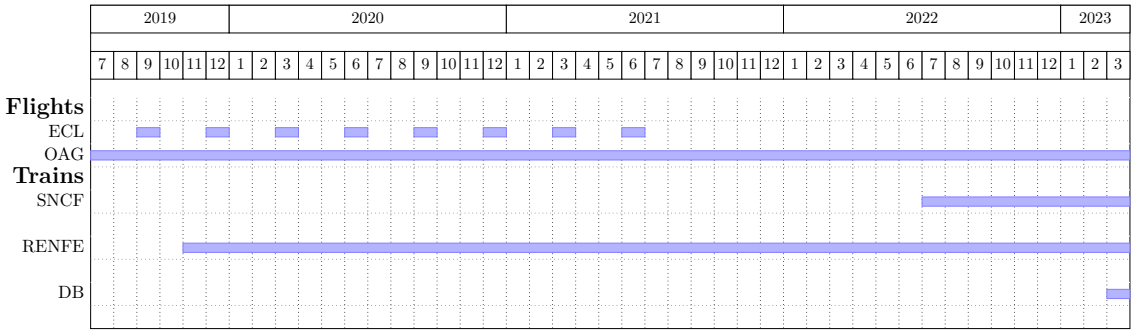


Figure 2.3 – Data sources and availability on March 2023. Blue lines correspond to dates when the data is accessible.

Madrid-Puerta de Atocha, the two main train stations of Madrid, and then transfer to the airport. For simplicity we limit the analysis to the Madrid-Puerta de Atocha train station. Rail data for these three train stations : CDG-HSR, FRA-HSR and Madrid-Puerta de Atocha are therefore collected. In the following, we focus on two periods : 2019 and 2021, to compare changes due to the COVID-19 outbreak. When historical data is available, the schedules of HSR across the three countries are collected on each website. Data from 2021 was collected at the time on the operator website. To retrieve data from 2019, French HSR schedules are downloaded from Transitland (Interline, 2022), and German HSR schedules are obtained from a git repository (Lockheed, 2021). Data temporal availability, according to each source, is summarised in Figure 2.3. In addition to schedule data, remark that transportation operators now propose API to access real-time schedule, denoted by GTFS-RT. In addition to schedule information, this data contains actual arrival and departure times for each train.

Data from December 2019 and September 2021 was extracted for each mode. These months are chosen since data on that period have been retrieved for the two transportation modes. In addition, gathering data from pre- and post-pandemic periods will allow us to compare the transportation system evolution after the COVID-19 outbreak.

2.1.2 Comparison and overview

Several features are essential to build timetables. In particular, we are interested in :

— *Carrier* : airline or railway operator in charge of the transportation. In case of airline alliance, there can be two carriers, the publish carrier, which sales the ticket, and the operating carrier, which effectively operates the flight ;

— *Origin* and *Destination* : origin and destination airports or train stations ;

— *Scheduled departure and arrival times* : off-block and in-block times ;

— *Actual departure and arrival times* : if the schedule corresponds to a past event, the flight or rail legs have been operated, and the true departure and arrival times are given. They may be different from the scheduled times in case of delay ;

— *Callsign/Tail number* : the aircraft or train identification. This number is unique and assigned to each vehicle (for instance aircraft "FHLVM"). This information is important because, if two successive flights are operated by the same aircraft, a sufficient time between the two flights should be planned in the timetable to ensure the unboarding of passengers, refuel, clean and boarding of new passengers. This minimum required time is referred to as the Minimum Turnaround Time (MTT) ;

— *Equipment* : it corresponds to the vehicle type (*e.g.* Airbus A320, Boeing 777, etc). This information is useful to determine the MTT between two flights since the turnaround time of a large aircraft is generally larger than for a smaller aircraft ;

— *Available seats* : number of passenger seats available ;

— *Passenger carried* : volume of passenger carried if the flight or the rail-leg already run (*i.e.*, historical schedule).

Subsequently, a processing step is initiated to extract and format the necessary features from each data source. The availability of each feature from raw data is presented in Table 2.1. For both air and rail data collected, the carrier, origin station (*i.e.*, airport or train station), destination station, scheduled departure time and scheduled arrival time information are available. In addition, Eurocontrol Data (ECL) for flight, and GTFS-RT for trains, provide actual departure and arrival times. Regarding vehicle information, only

Table 2.1 – Features availability for air and rail data sources

	Air		**Rail**	
Features	ECL	OAG	GTFS	GTFS-RT
Carrier	Yes	Yes	Yes	Yes
Origin	Yes	Yes	Yes	Yes
Destination	Yes	Yes	Yes	Yes
Scheduled departure time	Yes	Yes	Yes	Yes
Scheduled arrival time	Yes	Yes	Yes	Yes
Actual departure time	Yes	No	No	Yes
Actual arrival time	Yes	No	No	Yes
Vehicle callsign	Yes	No	No	No
Equipment	No	Yes	No	No
Available seats	No	Yes	No	No
Passenger carried	No	No	No	No

ECL share the aircraft callsign, and information on the equipment and seats are only

accessible using OAG Schedule Analyzer (OAG, 2023) data. Finally, no source provides the passenger volume carried in historical schedules.

As a summary, data from a week of September 2021 and a month of December 2019 are collected for trains in France, Germany and Spain. Note that flight schedules are shared in UTC time, whereas GTFS data is provided at the local time of the country. Figure 2.4 presents the number of high-speed trains scheduled for these countries for December 2019 and September 2021. These volumes are obtained after processing the GTFS files collected for these periods.

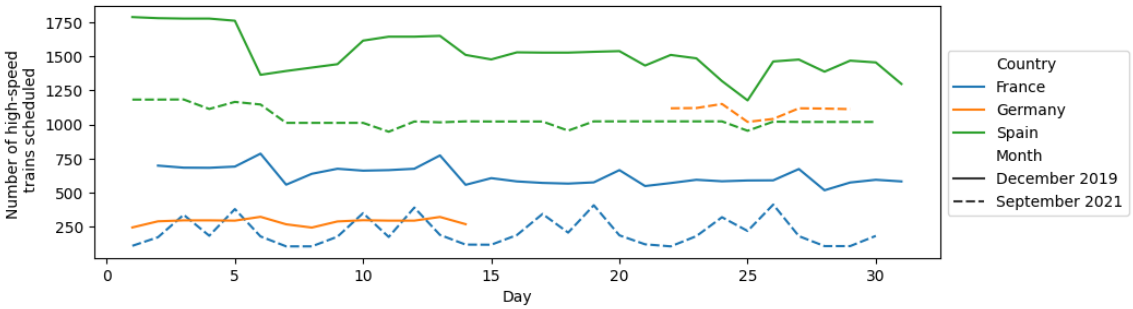


Figure 2.4 – Volume of long-distance trains scheduled in France, Germany and Spain in December 2019 and September 2021, according to GTFS data.

First, note that for Germany, as historical data is not available, there is no guarantee on the correctness of the data. An increase in the number of scheduled trains for Germany can be observed in 2021 compared with 2019, but this difference can be explained by the lack of information in the data of 2019. On the contrary, the number of scheduled long-distance trains in France and Spain reduced in 2021. Regarding absolute volumes, GTFS data from Spain includes HSR as in the French GTFS data, but also long and medium distance trains.

Regarding flight volumes, several thousands of flights are operated each day across Europe. The monthly volume of operated flights from June 2019 to September 2021, computed from Eurocontrol data, is presented in Figure 2.5. The COVID-19 outbreak significantly impacted the air transportation system. However, the year 2021 shows an increase in air traffic towards the pre-pandemic level. In September 2021, more than 600,000 flights were operated in Europe. Several thousands of flights and hundred of trains scheduled will therefore be handled in the dissertation.

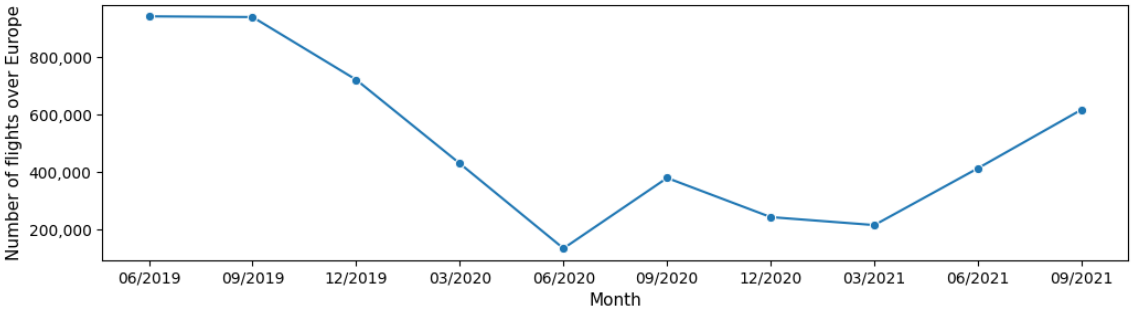


Figure 2.5 – Volume of commercial (passenger) flights operated over Europe for months of March, June, September and December from June 2019 to September 2021. Data is extracted from ECL.

2.1.3 Dealing with missing information

The previous sections highlight that important information is missing for timetable generation. Specifically, in Chapter 3, the developed synchronisation mechanism suggests making minor adjustments to existing schedules, assuming that aircraft and trains are already assigned to legs. Furthermore, it is more rational to synchronise train and flight schedules when they affect a significant number of passengers. For instance, coordinating a train with a specific flight is not relevant if no passenger transfer between the two. The objective of the multimodal demand simulation problem we are about to address is therefore to estimate such transfer volumes. This estimation relies on the passenger volume carried on each leg ; however, this information is never accessible from schedule data. In this section, two heuristics are proposed to retrieve both the vehicle assignment and the passenger volume carried in historical data.

2.1.3.1 Vehicle assignment

The train or aircraft assignment is an important information since operational constraints such as the minimum turnaround times will be considered afterwards. In order to simulate this information realistically, we propose a heuristic to assign a vehicle to each flight or train journey. To do so, two assumptions are made :

— each vehicle is owned by exactly one carrier,

— the number of available aircraft/train is not limited.

Figure 2.6 displays the distribution of the average turnaround time for each aircraft in June 2019 as a function of the in-flight time. One can observe that, generally, the turnaround time is above 30 minutes, and ranges between 100 and 150 minutes most of the time. In the following, we set the value of MTT to 30 minutes.

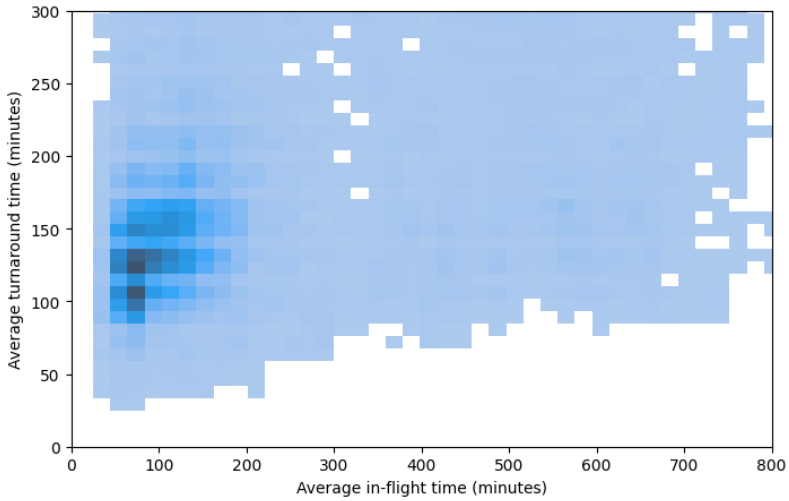


Figure 2.6 – Distribution of actual aircraft turnaround times for each day of June 2019 at CDG, FRA and MAD airports. A darker colour corresponds to a higher density.

Let L denote the index set of legs ordered by departure times and n denotes the number of legs ($n = |L|$). For each leg $l \in L$, L^l denotes the index set of legs that can be operated consecutively by the same vehicle of l, ordered by departure times. For instance, for a given flight l scheduled to arrive at CDG airport at 10am, the set L^l contains all flights scheduled to depart from CDG after 10am+MTT. Finally, for each leg $l \in L$, v_l denotes the vehicle ID that we are trying to determine, that operates leg l. As explained above, we assume that the number of vehicles available is not limited, but one can easily show that the number of legs is an upper bound for the number of vehicles required.

The principle of the heuristic we are proposing is presented by Algorithm 1 and is as follows. Its k^{th} iteration goes as follows. While there are legs without a vehicle assigned, the first element of L (*i.e.*, the earliest scheduled leg, denoted by i) is selected. The k^{th} vehicle is then assigned to operate leg i, and i is removed from L. If another leg can be operated after i by the same vehicle k, we assign k to this next leg, denoted by j. The same process is repeated for leg j and so on, until no more leg can be operated by vehicle k. A new iteration then starts and, again, the earliest remaining leg of L is selected, and a new vehicle $k + 1$ is assigned to it. The process is repeated until all legs are covered by a vehicle.

If additional flight information is available, the algorithm can be extended to improve the quality of the simulation. For instance, the airline information is accessible through OAG or ECL data. In addition, OAG data provides for each scheduled flight, the vehicle type and its seat configuration (for a same vehicle type, there are generally several possible seat configurations). For instance, an airline can have two aircraft of the same type, but

Algorithm 1 Vehicle assignment heuristic

Inputs :

L : Set of legs ordered by departure time

k_0 : First available aircraft ID

L^i : Set of legs that can be operated after leg $i \in L$, by a same vehicle

procedure VEHICLEASSIGNMENT(L, k_0, L^0, L^1, ..., L^{n-1})
- $A = [\,]$ ▷ Index list of legs with a vehicle assigned
- $V = [\,]$ ▷ List of vehicle identification assigned to each leg
- $k \leftarrow k_0$ ▷ Aircraft ID initialisation, $k \in \{0, 1, \ldots, |L|\}$
- **while** $L \neq \emptyset$ **do**
  - $i \leftarrow min\{l : l \in L\}$ ▷ Earliest departure flight
  - $v_i \leftarrow k$ ▷ Vehicle k is assigned to leg i
  - $A.append(v_i)$
  - $V.append(k)$
  - $L \leftarrow L \backslash \{i\}$
  - **while** $L^i \neq \emptyset$ **do** ▷ While legs can be operated by the same vehicle n
    - $j \leftarrow min\{l : l \in L^i\}$
    - $v_j \leftarrow k$
    - $A.append(v_j)$
    - $V.append(k)$
    - $L \leftarrow L \backslash \{j\}$
    - $i \leftarrow j$
  - **end while**
  - $k \leftarrow k + 1$ ▷ New vehicle
- **end while**
- **return** V, A, k

end procedure

the number of economic-class seats in one aircraft can be higher than in the other aircraft. Depending on the OD pair served and the distance travelled, some airlines can also adapt the seat configuration of their aircraft to maximise profit. The vehicle assignment procedure can therefore be applied to each subset of legs operated by the same airline, with the same vehicle type and seat configuration.

In order to assess the heuristic performance, a validation phase is made on flights of June 2019 at Madrid airport. For each day, Algorithm 1 is applied to OAG schedules, and as output the number of required aircraft to operate all flights is computed. This value is compared to the actual number of aircraft, obtained from Eurocontrol data. Results are presented in Figure 2.7. First, note that there is a difference between the number of flight schedules in ECL and OAG databases. The difference is on average of 0.8%. This can be

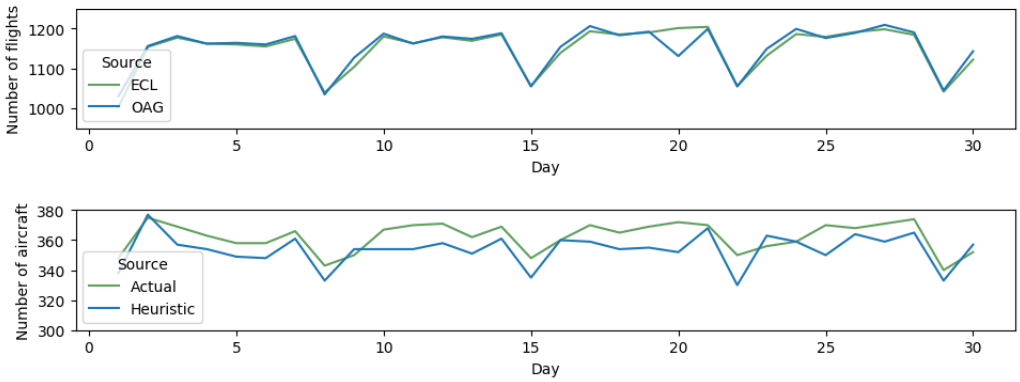


Figure 2.7 – The number of flights operated for each day of June 2019 at Madrid Barajas airport from Eurocontrol (green) and OAG (blue). Actual (green) and predicted (blue) number of aircraft required to operate these flights are displayed.

attributed to last-minute adjustments. Indeed, scheduled flights can be cancelled in practice or new ones can be added to accommodate disrupted passengers. Regarding the number of aircraft required, an average error of 2.6% is obtained for June 2019 at Madrid airport. The heuristic generally underestimates the number of aircraft. There are several reasons that can explain this underestimation. First, a low value of MTT is likely to reduce the number of aircraft required to operate a given number of flights. Indeed, with a large MTT, two flights scheduled at a reduced interval time cannot be operated by the same aircraft, an additional aircraft is therefore required. The value chosen of 30 minutes for MTT can thereby contribute to underestimate the number of aircraft. Including the flying time of flights operated (long/short range), or the airline type (regular/low cost) in the model could enhance the estimation of MTT and improve the results obtained. Another reason that can explain the underestimation of the number of aircraft is the fact that no information on the crew assignment and vehicle maintenance is available. Indeed, airlines should regularly schedule aircraft controls for safety purpose, increasing the number of required aircraft for the same number of scheduled flights. Finally, more aircraft than necessary may be planned to increase robustness, in case of disruptive events (aircraft breakdown, additional flight to schedule, etc.). These constraints, that we could not take into account here, certainly increase the number of aircraft required ; however, for simplicity, they will not be considered here.

Regarding train schedules, GTFS data does not provide vehicle assignment information. The same heuristic (Algorithm 1) is then employed to obtain a feasible timetable together with a vehicle assignment. Unfortunately, no database is accessible to validate our model on trains.

2.1.3.2 Volume of passenger carried

The historical volume of passenger carried on each flight is not accessible. Therefore, in this section, an estimation is proposed. Air passenger traffic volume to/from European airports are recorded in the Eurostat database (Eurostat, 2019). Values for the years 2019, 2020 and 2021 are collected. For each airport pair, with either the origin or destination (or both) within Europe, the monthly volume of passengers travelling between the two airports is given. Figure 2.8 displays the monthly volume of passengers carried within Europe from January 2019 to December 2021, recorded by Eurostat. Note that some values

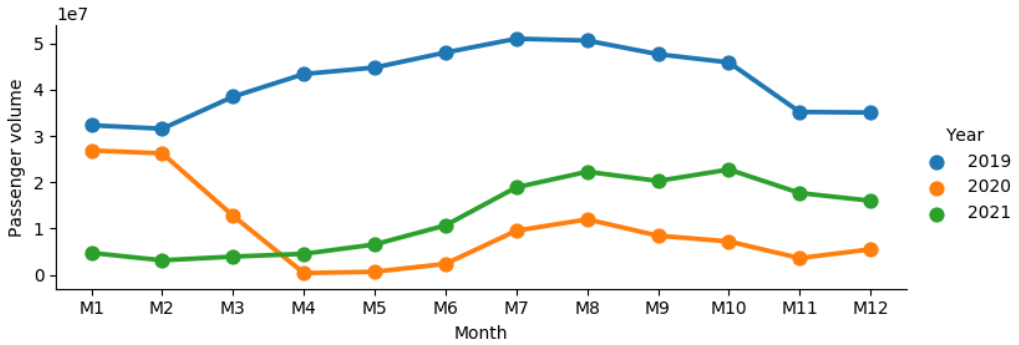


Figure 2.8 – Monthly volume of passengers carried within Europe from January 2019 to December 2021.

are equal to zero in the data, due either to missing information or to the fact that no passenger travelled on that route (especially during the COVID-19 outbreak). Table 2.2 summarises the characteristic of the data gathered. According to the International Civil Aviation Organization (2023), the volume of passengers carried in Europe in 2019 was above 1,171 millions, while according to Eurostat data, the volume of passengers recorded for 2019 was rather 504 millions. Such a difference can be explained by the lack of available data, putting light on the difficulty to find reliable and accurate data. For years 2020

Table 2.2 – Number of zeros and total number of passengers carried.

	2019	2020	2021
Zero values (%)	5.2	23.0	13.1
Total volume (millions of passenger)	504	116	151

and 2021, a higher share of zero values is observed. As explained above, this may be due to the COVID-19 lockdowns. Table 2.3 summarises the number of routes (OD pairs) served for several months from June 2019 to September 2021. These values are computed using ECL database. In September 2019, 40,152 OD pairs were served compared with

38,565 in September 2021. The pandemic reduces the number of route served, leading to no passengers on these historical routes in the Eurostat database. In addition, some of these zero values can be explained by the lack of available data.

Table 2.3 – Number of flight routes (OD pairs) served on the entire month, from June 2019 to September 2021. Values are computed from ECL.

Month	**Number of OD pairs served**
06/2019	41,299
09/2019	10,152
12/2019	30,582
03/2020	27,695
06/2020	19,173
09/2020	28,866
12/2020	22,933
03/2021	21,039
06/2021	33,302
09/2021	38,565

As no other passenger data is available, we choose to use these data as a basis to estimate the volume of passengers carried per flight, despite the above-mentioned missing information.

The procedure estimation we are proposing for a given day goes as follows :

1. for each OD airport pair, the daily passenger volume carried is estimated by dividing the monthly volume carried by the number of days ;
2. for each OD pair served on the day, the flight required frequency is then computed ;
3. finally, the volume of passengers carried per flight is estimated as the daily volume of passengers on a given OD pair, divided by the flight frequency.

If, for an OD pair, the volume of passengers carried is missing in the Eurostat data, we arbitrarily set the passenger volume to the average passenger volume per flight for the day, to avoid exogenous noise in the data.

This methodology is a simple approach to obtain a realistic estimation of the passenger volume carried. However, it has some drawbacks. First, the number of passengers per OD pair is estimated to be the same for each flight, which is probably not the case in practice. Indeed, the specific weekday and time of the day may influence the passenger demand. In addition, setting missing information to the average value of flying passengers does not change the macroscopic analysis, but this averaging may lead to large underestimations or overestimations of the number of passengers on some flights. In order to assess the quality of this primary approach, as no validation data is accessible, we used OAG flight

schedule data. Indeed, these data provides the number of seats available ; this information can therefore be compared to our estimation of the volume of passengers carried. Data from December 2019 is used since the average passenger load factors (*i.e.*, share of seats occupied) for 2020 and 2021 were lower than usual, due to the pandemic. For each flight f scheduled, the estimated number of passengers carried, denoted $\tilde{v}_f$, is compared with the available number of seats, s_f : for each flight, the difference $s_f - \tilde{v}_f$ is computed, and the distribution is displayed in Figure 2.9. One remarks that most of the volume estimated is lower than the number of available seats. Indeed, for some OD pairs, the monthly volume remains unknown. Setting the number of carried passengers to the average number of passengers carried per flight then underestimates the number of passengers, especially on long-distance flight (involving generally large numbers of passengers). This underestimation leads to differences sometimes larger than 400 passengers ! On the contrary, it appears that the number of passengers carried may be overestimated. In particular, on the OD pair Paris-Charles de Gaulle airport - Istanbul Sabiha Gokcen airport, the daily volume of passengers is estimated to more than 1400, while only two flights of 159 seats each are scheduled, leading to a very large difference value. Figure 2.10 displays the average difference between the number of seats available and the estimated volume. On average, the estimation is lower than the number of available seat by 43 passengers. If we compare with the average number of seats, this value corresponds to an average load factor of 75%. This value is only 6% under the load factor measured by IATA for December 2019 (International Air Transport Association, 2020). Finally, thanks to a collaboration with Aéroport de Paris (ADP), we had access to several months of traffic data collected at CDG airport. This data records

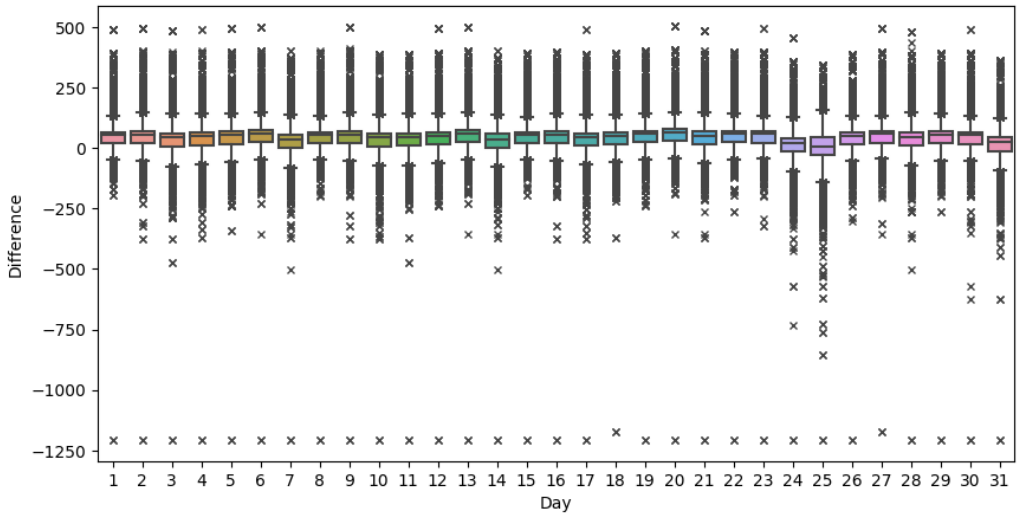


Figure 2.9 – Daily distribution of the difference between the available number of seats and the passenger volume estimated, for December 2019.

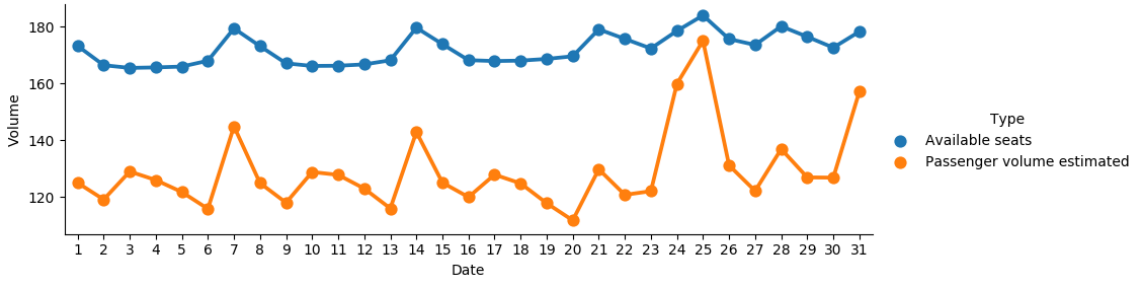


Figure 2.10 – Available number of seats (blue) and the passenger volume estimated (orange), for December 2019.

information on flights operated at CDG airport for several days of 2019. For each flight, the number of passengers carried is available. We therefore compare the number of passengers carried estimated with the actual volume of passengers carried for a week of December 2019. The distribution of absolute difference between the actual and estimated volumes of passengers carried, for each day, is displayed in Figure 2.11. The average difference in the

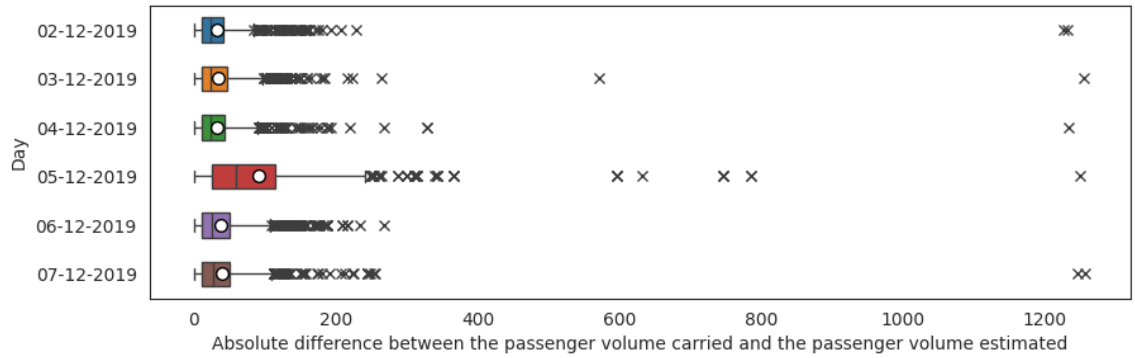


Figure 2.11 – Absolute difference (actual - estimated) volume of passengers carried, for flights arriving at and departing from CDG airport, for a week of December 2019. The daily average absolute difference is represented by a white circle.

number of passengers estimated ranges between 31 and 39 passengers for December 2, 3, 4 and 6. This value rises to 92 for December 5 due to a massive Air Traffic Controllers (ATC) strike that occurred that day in France, leading to flight cancellations, delays, re-booked passengers, etc. As we use monthly passenger volumes in our demand, the granularity of data does not allow us to capture such disruptive events, leading to worse results on these day. On average, over the six days tested, the absolute error in our estimation of the passenger volume carried is 35%.

Regarding train data, no information on passengers is available, in the remainder of the

thesis, an average value of 600 passengers is thereby considered (UIC-International Union of Railways, 2023b).

2.1.4 Data collection and standardisation conclusion

This section highlights one of the first challenges one faces when dealing with timetable coordination between multiple operators : data collection and standardisation. The existence of numerous data repository location, data prices, or diversity in data formatting make their re-use and integration a difficult task ; although they are necessary for further collaboration. The development of the GTFS format is a significant step towards unifying transportation data. While it is now widely adopted by ground transportation operators, its use remains limited for air transportation suppliers. Recently, the Spanish government pushed for transportation data accessibility through the MITMA project (Ministerio de Transportes, Movilidad y Agenda Urbana - NAP, 2023), and flight schedules within Spain are now published regularly under the GTFS format. This contributes to simplify the fusion of transportation data, paving the way towards large transportation data sharing.

Another issue that one must consider related to data is the lack of accurate available information, either due to privacy concerns or to lack of knowledge from transportation service providers. Therefore, this first section introduced two basic methodologies to retrieve vehicle assignment and the volume of passengers carried. These mechanisms rely on available data, and allows one to obtain realistic input for further optimisation processes.

The last missing information is the transferring demand between flights and trains. As it could not be obtained from any data at the moment, a data-driven methodology to simulate realistic instances is presented in the following section.

2.2 Air-Rail transfer demand estimation

One key information in transportation scheduling is the expected demand for a flight or a train. Indeed, as shown in Chapter 1, timetable generation and route planning rely on the expected revenue, which is based on the estimated number of passengers who will be carried. To date, no data measuring accurately the multimodal demand between trains and flights are freely accessible. To palliate this gap, this section introduces a data-driven passenger multimodal flow estimation. A brief literature review on passenger demand forecast is first presented. Then, the air-rail passenger flow simulation, based on constraint-programming, is proposed. Finally, the instances generated are presented and publicly released.

2.2.1 Literature review on passenger demand forecast

As transportation providers plan their service in order to maximise profit, having an accurate knowledge of future potential passengers' expectation is crucial to ensure profitability. Generally, the passenger demand is estimated by OD pairs. The OD demand

corresponds to the number of passengers who wish to travel from an origin O to destination D, in a given time period. Each step of the scheduling process is therefore guided by the expected profit, estimated according to the volume of passengers carried. Different levels of granularity are required depending on the scheduling step. As detailed by Banerjee *et al.* (2020), transportation demand estimation can therefore be divided into two categories :

— *macro* forecast, for strategic planning, such as network building or new infrastructure development ;

— *micro* forecast, related to day-to-day or within-day prediction, required for revenue management, schedule design or the management of real-time (tactical) operations.

In this dissertation, we shall focus on *micro* forecast, as our interest lies in building a daily timetable. *Micro* forecasting models can be divided into three main categories :

— time series models : predictions are based on historical observations (Wickham, 1995; Adrangi *et al.*, 2001; Dutta and Ghosh, 2012) ;

— causal models : explanatory variables, such as Gross Domestic Product (GDP), population size, distance between cities or travel time, are used to predict the passenger demand (Leng *et al.*, 2015; Ashiabor *et al.*, 2007; Cipriani *et al.*, 2014; Aston and Koopman, 2006) ;

— machine learning models (Tsai *et al.*, 2009; Srisaeng *et al.*, 2015).

Most transportation providers develop models that consider only one mean of transportation, including potential competition with other modes. A large number of studies analyse the impact of HSR on air travel demand and vice versa (Yang *et al.*, 2018; Clewlow *et al.*, 2014; Park and Ha, 2006). As for mono-modal travel, accurate multi-modal demand forecast is mandatory to capture which connections are the most important for passengers. Some studies have been carried out to estimate this demand in an urban area. For instance, Toqué *et al.* (2017) process smart card data to estimate the multimodal demand in Paris-La-Défense district area. The same technique is employed by Seaborn *et al.* (2009) and applied to London city. However, there is a noticeable gap in research addressing the estimation of air-rail transfer demand. Lewe *et al.* (2012) implement a dynamic system model to forecast the multimodal demand on the US transportation network. Their model is able to estimate the travel demand on air and ground transportation modes simultaneously as a function of exogenous parameters such as fuel price, population growth, or GDP. However, the ground transportation modes are aggregated under a general "ground mode" without distinction between car, rail or public transport. Li and Sheng (2016) analyse the passenger demand for air-rail integration. They attempt to measure the market share of three levels of integrated air-rail service, using passenger discrete choice model. More recently, Tan *et al.* (2022) propose a deep-learning approach to forecast rail-to-air transferring passengers. While these models can give insight on a global demand between air and rail, the exact number of passengers who transfer from a scheduled train to a scheduled flight, and vice versa, is not provided. Up to date, no passenger ground truth

data is available, that would allow transportation operators to follow passengers while they are not in a vehicle. Such information could be obtained through external sources, such as passenger-generated data. Indeed, these new data sources, generated by mobile phones, show a high potential to understand passengers behaviour. For instance, Marzuoli *et al.* (2019) use mobile network data of air passengers to measure US airports' catchment area. Crossing airport access and egress travels with ground transportation supply could bring new information of the transport modes used to access/egress the airport. García *et al.* (2016) also use mobile network data in Spain to infer the transport mode used for Madrid-Barcelona journeys. Lythgoe and Wardman (2002) estimate the demand for rail to/from British airports using such kind of data. Mobile phone data is also exploited to determine OD matrices. Alexander *et al.* (2015) leverage mobile phone data to extract the daily OD matrix within Boston area. Montero *et al.* (2019) process mobile phone data to retrieve the OD matrix and modal split in Barcelona. Burrieza-Galán *et al.* (2022) use mobile phone data to analyse passenger travel behaviour at Madrid-Barajas airport. Their methodology allows them to compare passenger travel patterns before and after the COVID-19 outbreak. Marzuoli *et al.* (2018) and Monmousseau *et al.* (2020) show that social network data such as Twitter data can provide accurate insight on the health of the air transportation system, especially under disruptive events. Another mobile data sources that may be used to trace passengers journey are wifi and bluetooth connection data (see for instance Aliari and Haghani (2012); Nikoue *et al.* (2015); Huang *et al.* (2019); Barcelö *et al.* (2010)).

Note that these data sets may have some limitations. First, not all passengers connect to the mobile network data (phone call, wifi hotspots or bluetooth) during their trips, and foreign travellers are generally not registered on local mobile phone networks. The volume of tracked passengers may therefore be underestimated. In addition, the connection to the network may be sparse, leading sometimes to large amount of time without information on the passenger location. The reconstruction of the trips should therefore include assumptions, generally resulting in estimations that can significantly differ from the actual passenger journeys. Finally, there are generally several mobile network service providers that operate on the same territory. Data from several operators should therefore be collected to capture information about all passengers. However, these data sources might be the only way to follow passengers during their door-to-door journey, and to estimate the passenger flow between scheduled trains and flights. Consequently, the demand for mobile phone data persists. Nevertheless, the acquisition and management of such data are frequently complex due to various factors, including cost, privacy concerns, and the infrastructure required to process this typically voluminous data.

As we do not have access to such data, we introduce here a data-driven methodology to model passenger flows between trains and flights at airports.

2.2.2 Constraint-programming approach

In this section, we propose a data-driven method to estimate passenger air-rail flows between airports and associated train stations. The problem to be addressed is first presented, followed by a proposition of a mathematical model. Then, the resolution approach is presented. In addition, a methodology to generate diverse solutions is detailed. Numerical results and real-case applications are finally presented in the last section.

2.2.2.1 Problem statement

In order to propose multimodal schedule synchronisation to passengers, it is crucial to understand passenger flows between modes. More precisely, we aim at estimating the volume of passengers who want to transfer from a train to a flight, at a given airport or within the same city area, and vice versa. We propose to address this problem by solving a routing problem on a transportation network graph. We set the spatial scope of the problem to a city area, composed of at least one airport and one train station, in which passenger transfers between trains and flights can occur. If the train station is located at the airport, the spatial scope corresponds to the airport. Figure 2.12 illustrates the possible passenger transfers considered. In this figure, each arrow represents a particular type of transfers that passengers can make within the area : travel directly from their home location to the train station or the airport, connect with a train or a flight from another leg, or exit the transportation system (airport or train station) to reach their final destination.

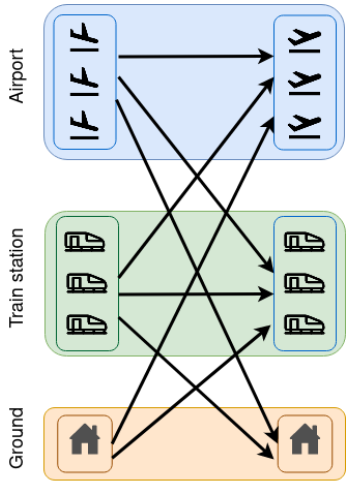


Figure 2.12 – Passenger transfers within a city area.

The objective consists in estimating how many passengers transfer between each leg-pairs.

2.2.2.2 Mathematical formulation

We propose to model the problem as an Integer Linear Programming (ILP) problem. The required sets, input data, decision variables and constraints are presented in this section.

2.2.2.2.1 Set definitions The transportation network is represented by a graph $G = \{V, E\}$, where the vertex set, V, is partitioned into the arriving-leg vertex set, V^A, and the departing-leg vertex set, V^D, and where the set of arcs, E, corresponds to *feasible* transfers between legs. A transfer is considered as *feasible* if the departure time of the second leg is after the arrival time of the first leg. In this sense, we impose a lower temporal limit, denoted Minimum Connection Time (MCT), to allow passengers to have sufficient time to transfer. Similarly, we assume that above a duration threshold, the connection time (time difference between the departure time of the second leg and the arrival time of the first leg) is too long for passengers to consider the connection. Therefore, we define a Maximum Acceptable Connection Time (MACT), above which the transfer is considered as "unfeasible". These values differ depending on the connection type considered. For instance, for passengers transferring from a flight to a train, the MCT is lower than for passengers transferring from a train to a flight, since there is no train station security screening process. In addition to temporal feasibility, for realism purpose, we limit the set of connecting legs to *relevant* ones. More precisely, we assume that passengers connect with a leg that is not returning to the passenger origin city. Additionally, we assume that if there exists a direct flight or train between two cities, passengers will choose the direct connection. For instance, when travelling from Lille to Marseille, we assume passengers will either use a high-speed train directly from Lille to Marseille or take a direct flight between the two cities. Rail-air connections for that OD pair at CDG airport are then not considered, as they would substantially increase the total travel time. A pre-processing step is therefore run to filter out relevant connections.

Let m denote the number of arrivals and n the number of departures. We assume that the station is served by a set of transportation modes (*e.g.*, air, rail, bus, car, subway) represented by the index set K. The set of arrival vertices is further partitioned as $V^A = \bigcup_{k \in K} V_k^A$, where V_k^A is the index set of arriving legs of mode k, $k \in K$. The set of departing legs is similarly defined : $V^D = \bigcup_{k \in K} V_k^d$, where V_k^D is the index set of departing legs of mode k, $k \in K$. The vertex set also includes ground nodes, v_0 and v_{m+1}, to represent *outbound* passengers, *i.e.*, passengers arriving from the city, and *inbound* passengers, departing to the city, respectively. Similarly, the edge set includes transfer edges from each arriving train and flight to the ground node v_{m+1}, and transfer edges from the ground node v_0 to each departing train and flight. Figure 2.13 illustrates the notations introduced.

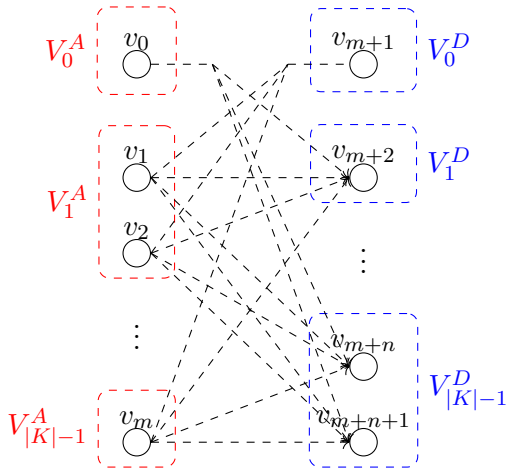


Figure 2.13 – Graph representing a transportation station. Nodes corresponds to arriving and departing legs, arcs to feasible transfers between them. Nodes are aggregated per means of transportation.

2.2.2.2.2 Input data For each leg node $v_i, i \in V$, the volume of passenger carried is denoted w_i. At the station, arriving passengers have two options. They can either transfer to another leg, that could be within the same mode or not, or leave the station. Similarly for departing passengers : they can either arrive from another mode or directly arrive from the ground. The total volume of departing passengers can be computed as $W^{\text{out}} = \sum_{i \in V^D} w_i$. In particular, we can compute the total passenger volume leaving the station with mode k as $W_k^{\text{out}} = \sum_{i \in V_k^D} w_i$. Consider a pair of modes $(k, k') \in K \times K$, let $p_{kk'}$ be the share of passenger volume that transfers from mode k to k' at the station. We compute the total volume of passengers transferring from mode k to k' at the station, $(k, k') \in K \times K$, as follows : $W_{kk'} = p_{kk'} W_{k'}^{\text{out}}$. For each arriving or departing leg node $v_i, i \in V$, the maximum volume of transferring passengers, to or from mode $k \in K$, depends on the number of feasible connections. For instance, for the last arriving flight of the day, passengers have no transferring option. We consider then that they all leave the station and the maximum passenger volume that transfers to a flight is zero. In practice, passengers may connect with a flight or a train the following day, but this is not considered in the present study. For each arriving leg node $v_i, i \in V^A$, we define V_{ki}^D to be the set of departing-leg nodes of mode k with which passengers can make a connection. Similarly, for each departing-leg node $v_i, i \in V^D$, we define V_{ki}^A to be the set of arriving-leg nodes of mode k from which passengers of v_i can arrive. For each leg node v_i, $i \in V$, $\overline{y_{ik}}$ denotes the maximum volume

of passengers who can transfer with a leg of mode $k \in K$. It is computed as follows :

$$\overline{y_{ik}} = \begin{cases} \min(w_i, \sum\limits_{j \in V_{ki}^D} w_j), & \text{if } i \in V^A, \\ \min(w_i, \sum\limits_{j \in V_{ki}^A} w_j), & \text{if } i \in V^D. \end{cases}$$

2.2.2.2.3 Decision variables The objective consists in determining the volume of transferring passengers between each legs. For each $i \in V^A$ and $j \in V^D$, we define the primary integer decision variable x_{ij} to be the volume of passengers who transfer between leg v_i and v_j. For the sake of clarity, we also introduce for each leg $v_i, i \in V$, and for each transportation mode $k \in K$, the auxiliary integer decision variable y_{ik} that corresponds to the volume of passengers transferring from leg v_i to/from a leg of mode k.

2.2.2.2.4 Constraints An acceptable solution must satisfy the following constraints :

$$\sum_{j \in V_k^D} x_{ij} = y_{ik} \qquad k \in K, i \in V^A, \tag{2.1a}$$

$$\sum_{i \in V_k^A} x_{ij} = y_{jk} \qquad k \in K, j \in V^D, \tag{2.1b}$$

$$\sum_{k \in K} y_{ik} = w_i \qquad i \in V, \tag{2.1c}$$

$$y_{ik} \leq \overline{y_{ik}} \qquad k \in K, i \in V, \tag{2.1d}$$

$$\sum_{i \in V_k^A} y_{ik'} = W_{kk'} \qquad (k, k') \in K \times K, \tag{2.1e}$$

$$\sum_{j \in V_{k'}^D} y_{jk} = W_{kk'} \qquad (k, k') \in K \times K, \tag{2.1f}$$

$$x_{ij} \in \{0, 1, \ldots, \min(w_i, w_j)\} \qquad i \in V^A, j \in V^D, \tag{2.1g}$$

$$y_{ik} \in \{0, 1, \ldots, \overline{y_{ik}}\} \qquad i \in V, k \in K. \tag{2.1h}$$

Constraints (2.1a) and (2.1b) correspond to classical flow constraints. Constraints (2.1c) ensure that the number of passengers assigned to each leg equals the passenger volume carried. Constraints (2.1d) stipulate that the share of transferring passengers per leg to/from each mode does not exceed the maximum authorised limit. Constraints (2.1e) and (2.1f) ensure that all passengers are assigned. Finally, constraints (2.1g) and (2.1h) specify the definition domain of the decision variables. A particularity of this optimisation model is that there is no objective function. It is a feasibility model : any feasible solution provides one possible passenger transfer flow within this spatial area.

2.2.2.3 Resolution approach

Since no objective function is defined, one can easily solve the optimisation problem by minimising any constant objective function : a linear-programming solver will return the first feasible solution found. However, in our application context, one issue remains with solutions obtained via a state-of-the-art linear programming solver : it will return a solution with saturated constraints (the simplex method restricts its search to extreme points of the feasible-domain polyhedron). For instance, if 100 passengers arrive by a train at 1pm and a flight is scheduled to depart at 4pm, the linear solver is likely to return a solution with the 100 passengers transferring to the same flight if the number of seats is sufficiently large. The probability that such an event occurs in practice is really low. Therefore, in order to simulate more realistic passenger flows, we rather choose to solve our feasibility problem with Constraint-Programming (CP) (Rossi *et al.*, 2006). CP is based on the following backtracking principle :

— at each iteration, a decision variable is selected and assigned to a value of its definition domain ;

— the definition domains of associated variables are reduced (*constraint propagation*), resulting from this assignment ;

— while no *inconsistency* is found (*i.e.*, constraint violation) the tree search continues.

To speed up the resolution, we set the choice of the decision variable at each iteration based on the so-called *weighted constraint algorithm* (Boussemart *et al.*, 2004). The objective of this algorithm is to detect the hard parts of the problem by finding the constraints that are the most often violated during the tree search process : each time a constraint is violated, its weight is increased. The algorithm then proposes to assign in priority variables that are involved in the large-weight (hard) constraints, in order to concentrate first on the hard parts of the problem. Finally, in order to avoid obtaining solutions involving too numerous saturated constraints, the value assignment is randomly (following a uniform probability law) made among what remains of the definition domain of the decision variables.

2.2.2.3.1 Diverse solution set generation In order to simulate a sufficiently-diverse instance set representing several different days, we aim at increasing the diversity of the generated solutions. In the following, let X^j denote the j^{th} solution vector obtained, and x_i^j the value of variable i in solution X^j. Several studies propose to generate diverse solution by maximising a distance from each of the solutions previously found. This problem is known as the DIVERSEkSET problem. Its objective consists in finding k solutions as diverse as possible. For this, several distance metrics can be used :

1. The Hamming distance : $\sum_i \mathbb{1}_{x_i^j \neq x_i^{j'}}$ computes the number of differences in the variables of two solutions (Hebrard *et al.*, 2005; Nadel, 2011; Nguyen *et al.*, 2012) ;

2. The Manhattan distance : $\sum_i |x_i^j - x_i^{j'}|$ computes the sum of absolute difference in the variables of two solutions ;
3. The Euclidean distance : $\sum_i (x_i^j - x_i^{j'})^2$ computes the l_2 norm of the vector $X^j - X^{j'}$.

However, in our application context, maximising any of these metrics leads to major issues. First, the use of the Hamming distance would be relevant for dealing with a vector X composed of binary variables, but in our case, the variables are integers. For instance, changing all variable values by only one passenger maximises the Hamming distance, although the solution obtained would then not be significantly different from the original solution. Second, maximising the Euclidean distance is likely to lead to solutions with numerous saturated constraints (some connections completely filled with passengers, and no passenger, or very few, on the remaining ones). This would generate passenger flow scenarios that are not representative. We therefore decide to use the Manhattan distance for our application. However, maximising a convex piecewise linear function requires extra binary variables resulting thereby in an Integer Linear Programming problem, known to be difficult.

We therefore propose the following methodology :

— let K be the number of solutions searched, and let M_0 be some-threshold Manhattan distance set by the user ;

— for each step k, $k = 1, 2, \ldots, K$, we add to the feasibility problem a constraint stipulating that the new solution should lie at least at a Manhattan distance M_0 of each of the previous solutions found $(X^1, X^2, \ldots, X^{k-1})$;

— if no such solution can be found, the value M_0 is iteratively decreased by a constant C such as

$$M_n = M_0 - nC, \tag{2.2}$$

$n = 1, 2, 3, \ldots$ until a solution is found or $M_n \leq 0$.

Numerical results are presented in the following section.

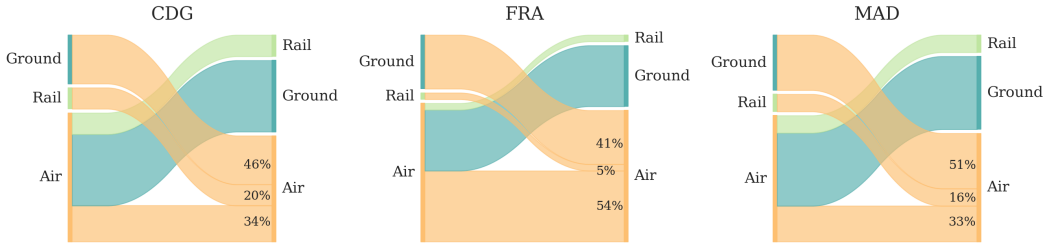
2.2.3 Numerical results

The methodology proposed to generate a diverse instance set is tested on the three hubs already mentioned : CDG, FRA and MAD airports. Subsection 2.2.3.1 details the collection of airport modal shares and the estimation of global transferring passenger volume. Subsection 2.2.3.2 presents the generated instance set.

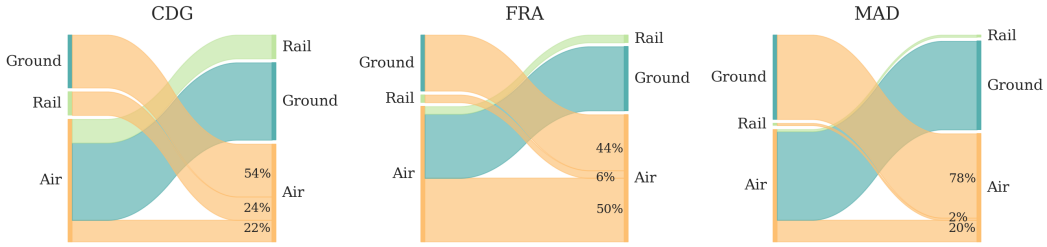
2.2.3.1 Modal shares and transferring passenger volumes

For Frankfurt airport, the rail and air transfer shares, for years 2019 and 2021, are both obtained from the airport website Fraport (2021). Regarding Madrid-Barajas and Paris-Charles de Gaulle airports, the air transfer rates of 2019 are obtained using the values computed by Maertens *et al.* (2020). Since no data is available for 2019, values of 2018

are used. The air transfer share value for CDG airport for 2021 is collected from Aéroport de Paris financial report (Groupe ADP, 2021). For MAD airport, data from the study of Burrieza-Galán *et al.* (2022) is used for both air and rail transfer rates of 2021. The values estimated for 2020 are used since no information for 2021 is available. Finally, the rail share at CDG airport is estimated using SNCF open data (SNCF, 2023a), as the annual number of passengers stopping at each train station is available. This value has been recorded for the CDG-HSR train station. For instance, in 2019, it corresponds to 20% of the total passenger volume departing from CDG. Values for 2019 and 2021 are therefore used as rail share estimators for the airport. All these values are summarised in Figure 2.14. One



(a) December 2^{nd}, 2019



(b) September 22^{nd}, 2021

Figure 2.14 – Modal share of CDG, FRA and MAD airports on December 2^{nd}, 2019 (top) and September 22^{nd}, 2021 (bottom).

observes that the three hubs have different characteristics. FRA is the airport with the highest air passenger transfer rate. In 2019, on average, 54% of passengers departing from FRA airport arrive from a connecting flight. Regarding air-rail connections, CDG is the one with the highest multimodal transfer rate (20% for CDG, 5% for FRA and 16% for MAD in 2019). From these values, the input data $W_{k,k'}$ is estimated for each mode pair (k, k') at these airports by multiplying the total departing air passenger volume with the corresponding modal share.

2.2.3.2 Instance set generated

Computations are performed on a laptop equipped with an AMD Ryzen 5 4500U CPU and 16 GB RAM. The feasibility problems are solved with the open-access CP solver Choco-solver (Prud'homme and Fages, 2022). For each iteration, we set a time limit of 60 seconds. An iteration n corresponds to the search of a solution that is at least at Manhattan distance M_n from the previous solution (see Equation 2.2). The multimodal passenger flow simulation is tested on the three hubs : Paris-Charles de Gaulle airport (CDG), Frankfurt airport (FRA) and Madrid-Barajas airport (MAD). For each airport, K=10 solutions are computed. The initial Manhattan distance M_0 is arbitrarily set to five times the number of decision variables. This corresponds to an average difference of five passengers per connection between each solution. The multiplying factor C used to reduce the required Manhattan distance is set to 10,000. Finally, in order to obtain realistic solutions, we limit the maximum number of transferring passengers between two given legs. Preliminary results suggest to set to 30 the maximum number of passengers allowed to transfer from a flight to another flight, and to fix to 60 the maximum number of passengers transferring from a flight to a train (and vice-versa). Using lower values for these parameters generally yields unfeasible problems. Table 2.4 summarises the number of primary decision variables (x_{ij}), and the computation times for each airport and each day considered in 2019.

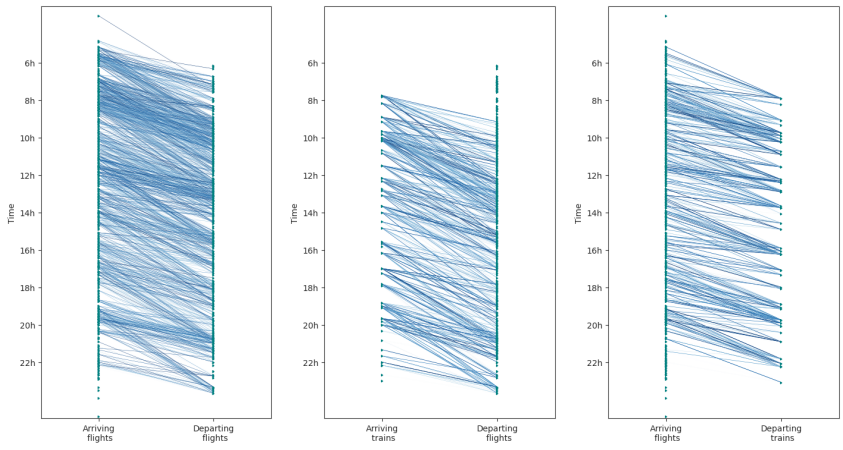
The number of primary decision variables is higher for FRA airport, since the air transfer rate is the highest one. This explains computation times that are on average larger than for the other two airports. The large computation times for instance 2 corresponds to the iterative decrease of the prescribed Manhattan distance M_0. Indeed, at the second step, no solution that satisfies the diversity constraint is found in less than one minute, yielding sometimes several successive decreases of the value of M.

Figure 2.15 displays a representation of two instances of CDG airport. Nodes correspond to arriving or departing flights and trains. An arc is drawn when there are passengers transferring between the two legs, and the arc width is related to the passenger volume.

All these instances and the associated schedules computed in this section are released and available from Buire (2023). This simulated passenger demand data can serve as realistic inputs to test or compare synchronisation mechanisms between air and rail.

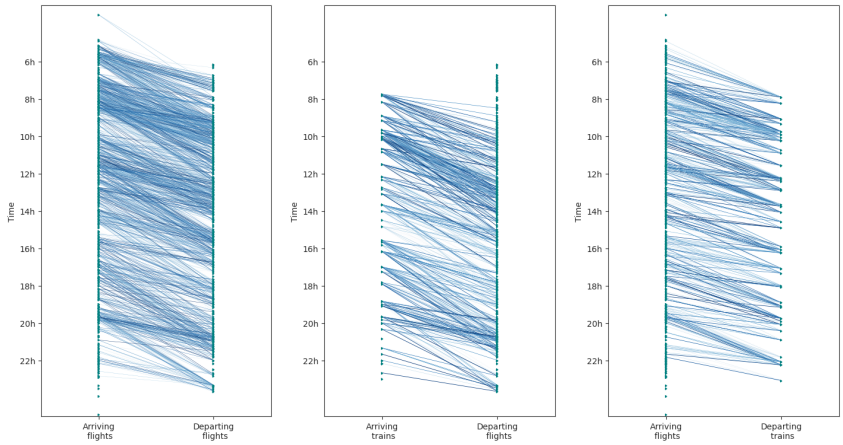
Table 2.4 – Number of decision variables and computation time (in seconds) of the passenger demand simulation for a week of December 2019.

Date	**02-12-2019**			**03-12-2019**			**04-12-2019**		
	CDG	**FRA**	**MAD**	**CDG**	**FRA**	**MAD**	**CDG**	**FRA**	**MAD**
# variables	27,495	40,335	23,670	25,674	40,151	21,833	27,358	39,288	21,510
Instance									
1	4.28	7.84	4.98	4.18	7.16	4.39	4.24	6.60	4.17
2	248.21	919.37	307.39	185.17	919.25	246.65	247.81	855.46	246.46
3	5.84	13.74	6.46	4.76	12.19	5.63	6.21	9.19	4.64
4	5.84	11.7	5.41	5.54	12.05	5.71	5.83	8.99	4.10
5	6.46	9.09	6.46	4.55	13.25	7.41	5.89	9.29	5.27
6	5.62	10.15	6.83	4.65	10.80	5.38	6.13	7.61	4.91
7	6.54	11.32	6.10	5.06	11.05	6.33	6.78	8.23	5.05
8	6.27	11.61	5.80	5.12	11.67	6.00	6.34	8.87	5.08
9	7.3	12.8	5.56	5.67	11.59	6.61	7.83	9.33	5.34
10	8.46	13.67	6.57	5.95	13.82	7.41	8.95	10.65	5.64
Average	30.482	102.129	36.156	23.065	102.283	30.152	30.601	93.422	29.066
Date	**05-12-2019**			**06-12-2019**			**07-12-2019**		
	CDG	**FRA**	**MAD**	**CDG**	**FRA**	**MAD**	**CDG**	**FRA**	**MAD**
# variables	14,450	34,230	17,127	23,195	39,136	10,005	24,628	37,228	9,882
Instance									
1	1.89	6.31	3.32	4.08	6.95	1.33	4.25	6.35	1.3
2	2.08	672.44	64.85	64.51	855.09	1.32	124.89	796.65	1.41
3	2.48	10.6	4.69	65.06	8.06	1.48	5.38	11.51	1.13
4	2.18	11.83	4.18	6.15	9.47	1.4	4.88	11.05	1.58
5	2.83	9.71	64.48	5.71	10.57	1.35	5.3	9.09	1.34
6	2.26	8.58	4.42	5.8	9.19	1.79	65.14	9.05	1.91
7	3.26	9.34	4.5	5.56	9.79	1.91	5.1	10.45	1.77
8	3.72	9.84	4.39	5.99	10.74	1.55	5.46	10.45	2.39
9	3.53	10.3	4.09	6.7	10.66	2.12	5.42	11.21	2.46
10	3.36	11.95	4.73	6.14	12.36	2.16	7.23	12.7	2.61
Average	2.759	76.09	16.365	17.57	94.288	1.641	23.305	88.851	1.79



(a1) Flight-to-flight **(a2)** Train-to-flight **(a3)** Flight-to-train

(a) Instance 1



(b1) Flight-to-flight **(b2)** Train-to-flight **(b3)** Flight-to-train

(b) Instance 5

Figure 2.15 – Two passenger flow simulations between flights obtained with CP, at CDG airport. The width of the arc is related to the number of passengers making that connection.

2.3 Conclusion

This chapter summarises the data processing steps that we proposed to gather, clean and standardise transportation supply data. The overall process is depicted in Figure 2.1. Gathering recent data, at a the European scale, and under a standard format remains a challenge. This section highlights the difficult and laborious task to collect recent data. Historical schedules are not always stored, and if so they are only accessible at a large monetary cost. To overcome this issue, rail schedules should be collected from each operator website, often in the national language. In addition to temporal discrepancies, there is no standard format to publicly share the air transportation data. Progresses have been achieved regarding ground public transportation data with the creation of the GTFS format and the recent incentive to publicly share ground transportation data, but airlines remain reluctant to publish largely their data.

In addition, this chapter proposes and validates heuristics allowing users to retrieve callsign information and passenger volume carried. Finally, a data-driven multimodal passenger flow simulation is introduced. Instances are generated on three major European hubs : Paris-Charles de Gaulle, Frankfurt and Madrid-Barajas airports. Theses instances are publicly released, allowing transportation schedulers to test further and compare schedule synchronisation algorithms on realistic instances.

In this thesis, estimation is performed on historical schedules, since they allow us to validate our methodologies. However, the estimation methodologies we introduced can also be applied to address missing information issues for future schedules (vehicle assignment and passenger transfer volume). Indeed, airlines and railway operators can accurately estimate the number of passengers who will be carried in flights and trains several days or weeks before operations, through ticket sale information. In the case of longer term planning, airlines and railway operators forecast can also bring insight on the expected demand for each flight or train. Our multimodal passenger demand estimation model can therefore be applied to the forecasted volume of passengers and not only to historical data. While this information is not publicly accessible, one can assume that in the case of collaboration, each transportation stakeholder would benefit from sharing information with others.

Chapitre 3

Air-rail timetable synchronisation minimising the impact on existing schedules

In this chapter, two mathematical optimisation models are introduced for timetable synchronisation and resolution tools are developed to improve air and rail connectivity at airports, in terms of connection time. As illustrated in the literature, one of the key challenges in air-rail synchronisation is managing the duration of the passenger connections. This chapter, divided into two sections, aims at proposing small schedule adjustments for both air and rail transportation service providers, that are beneficial for passengers. Section 3.1 introduces schedule synchronisation model at a strategic level, based on a connectivity metric. The methodology is applied to an air-rail transportation network across three European countries : France, Spain and Germany, and the potential benefits for passengers are assessed. Section 3.2 presents an adaptation of the proposed model to real-time situations application. The methodology aims at rescheduling trains and flights to mitigate the impact of a disruption on the air or the rail transportation networks, leading to significant delays and inducing potential missed connections for passengers. The objective of our model is therefore to reschedule flights and trains so as to wait for the delayed passengers.

3.1 Strategic planning horizon

The objective of this section is to propose a model that encourages passengers to use a combination of trains and flights in their trip. To this end, we introduce a methodology to "improve" passenger multimodal transfers at the expense of a limited cost for transportation service providers. The methodology is assumed to be applied several weeks in advance, in order to encourage passengers to book multimodal connections. Before diving into optimisation and synchronisation, it is first necessary to define what is meant for a

train and a flight to be "synchronised". To that aim, a first synchronisation metric is introduced and used as a basis for the optimisation model that we shall introduce. Then, the mathematical optimisation model for timetable synchronisation is proposed.

The methodology is then tested on the Western European transport network, using the simulated passenger-flow instances computed in the previous chapter.

3.1.1 Assessing passenger multimodal connection comfort

When passengers travel, they can make direct connections between their origin and destination if there is a direct flight or train. Otherwise, they have to transfer between two legs within the same network (*intramodal* connection), or they can shift to another mode (*intermodal* connection). In both cases, a sufficient amount of time must be planned to let passengers transfer from one leg to the next one. In the following, we call *connection time* between a first leg l_1 and a second leg l_2, the time difference between the scheduled departure time of leg l_2 and the scheduled arrival time of leg l_1. Intramodal connections are generally made within the same station. This is particularly true for the air transportation system where transfers take place at hub airports. In this case, the connecting time mainly consists of *wait* time at the station. However, sometimes, passengers may have to transfer between two different stations. Some air or rail connections involve changing stations and for intermodal connections, the train station may not be located at the airport. In this case, passengers have to follow several steps :

— Leave the first station : passengers get off the first vehicle and collect their luggage. This time is referred to as the *arriving processing time*.

— Transfer to the next station : passengers use a third means of transport to reach the next airport or railway station. This time is denoted *transfer time*.

— When passengers arrive at the next station, if it is an airport, they must check in, drop off their luggage, and go through the airport security process. This time is called *departure processing time*.

— Finally, they wait for the next leg of the journey. This corresponds to the *wait time*.

In order to make a connection feasible, passengers need a minimum amount of time between legs. This time should include the transfer time between the two stations in the case of multimodal connections (if the airport is not directly linked with an HSR station for instance), and possibly a station processing time (check-in, security screening at airport, border control, baggage collection after a flight, etc.). Recall that we call Minimum Connection Time (MCT) this minimum connection time, *i.e.,* the minimum time required to connect between two legs if there is no interruption during the connection process. More precisely, the MCT value corresponds to an optimistic situation in which the entire station process is seamless. For instance, abnormally-long queuing at check-in counters, security checks and border controls, or random searches at the security screening are considered as disruptions. Similarly, the Maximum Acceptable Connection Time (MACT) denotes the

connection duration above which the connection is considered as too long for passengers (such a connection is unlikely to be booked by passengers due to excessive wait time).

We propose to evaluate the quality of a connection time with a piecewise-linear cost function c, representing the passenger connecting disutility as a function of the connection time t between a flight and a train. Figure 3.1 illustrates such a cost function. More precisely, t corresponds to the duration between the departure time of the second leg, and the arrival time of the first leg. We define a connection time interval $[\underline{t},\overline{t}]$ that contains transfer durations that are considered as *suitable* for passengers. This assumption is similar to that of Theis *et al.* (2006) who show that passengers do not necessarily try to minimise their connection time but they rather look for a suitable connection time. In the remainder of this dissertation, a *short* connection refers to a connection whose duration lasts between MCT and $\underline{t}$. Similarly, connections with a duration between $\overline{t}$ and MACT are referred to as *long* connections.

In practice, the bound values chosen, MCT, MACT, $\underline{t}$ and $\overline{t}$, depend on the connection type considered. This allows one to take into account a longer connection time, for example, for passengers catching a flight with an international destination, or when the train station is not located at the airport. In the sequel, let Θ denote the index set of the possible connection types (from a train to a domestic flight, from a train to a non-domestic flight, from a flight to another flight, or from a flight to another train), and MCT_θ, MACT_θ, $\underline{t_\theta}$ and $\overline{t_\theta}$ denote the parameters defined for connection type θ, $\theta \in \Theta$. For simplification, we consider in this study a cost function involving only two breakpoints, $\underline{t_\theta}$ and $\overline{t_\theta}$. However, the model can easily be adapted to several breakpoints, increasing the unit cost of very short and very long connections. One can easily adapt such a model featuring additional

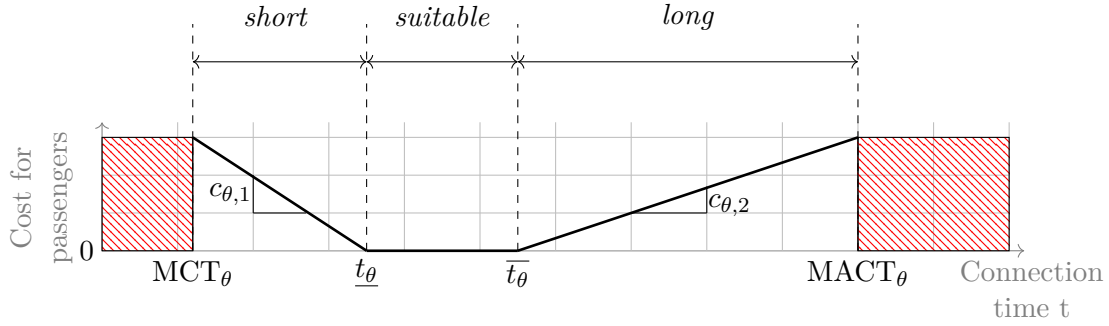


Figure 3.1 – Passenger disutility c_θ for a connection type θ as a function of the connection time t.

breakpoints as this does not affect the computational complexity of our model (minimising convex piecewise linear functions boils down to linear programming). It is reasonable to assume that short connection times are less desirable than long connection times. Indeed, short connections are not robust for passengers in case of delay : passengers are likely to

prefer connections with an additional buffer time to limit the risk of missed connections. Consequently, the unit cost of short connection times is assumed to be higher than long ones. The explicit formula for the passenger disutility function, as illustrated on Figure 3.1, is the following one :

$$c_\theta(t) = \begin{cases} c_{\theta,1}(t - \underline{t_\theta}) & \text{MCT}_\theta \leq t \leq \underline{t_\theta} \\ c_{\theta,2}(t - \overline{t_\theta}) & \overline{t_\theta} \leq t \leq \text{MACT}_\theta \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

As the connection time between two legs gets closer to the suitable interval $[\underline{t_\theta}, \overline{t_\theta}]$, the cost of the connection decreases linearly towards zero. Remark that this value zero is an arbitrary constant. Since connection time is to be optimised in the next sections, this value zero can be set to any other constant C, defined by the user. Indeed, minimising $f(x)$ and minimising $f(x) + n_t.C$ (where n_t is the total number of transferring passengers, which is another constant) will both lead to the same optimal solution. The objective of the air-rail strategic timetable synchronisation problem is to adjust timetables in order to ensure smooth passenger transfers. In other words, the aim is to create an integrated timetable with connection times ranging between MCT_θ and MACT_θ, and a cost, c_θ, that is as low as possible for each passenger having a connection of type $\theta \in \Theta$.

A timetable synchronisation model at the strategic planning horizon (several weeks before operation) is proposed in the following section.

3.1.2 Problem definition and mathematical formulation

When passengers want to use a train to reach the airport, for instance, they look for a train that is scheduled to arrive well before the departure time of their flight, so that they can transfer comfortably. Today, flight and train schedules are not designed to provide passengers with these smooth options. The question is therefore the following one : *is it possible, with limited changes to established schedules, to improve passenger multimodal connections, without impacting intramodal connections ?* A first model was proposed by Buire *et al.* (2022) to synchronise train and flight schedules at a single airport level. However, this model has two major drawbacks. First, air connecting passengers are not considered, while they are the priority for airlines. Second, network effects are not taken into account. Indeed, changing the departure time of a train or an aircraft at a station will automatically change its arrival time at the next station served.

Here, we propose a model that addresses these two previous limitations. Subsections 3.1.2.1 and 3.1.2.2 introduce the multimodal transportation network model and the passenger connection model. Subsection 3.1.2.3 presents the operational constraints considered in this thesis and how they are integrated in our optimisation problem. Finally, Subsections 3.1.2.4 and 3.1.2.5 present the mathematical formulation of the optimisation problem and the resolution approach chosen.

3.1.2.1 Multimodal transportation network model

In the following, index sets related to the air transportation network and index sets related to the rail transportation network will be super-scripted by "air" and "rail", respectively. The index set of stations (nodes) of the considered network, noted $\mathcal{N}$, is partitioned into the airport index set, $\mathcal{N}^{\text{air}}$, and the train station index set, $\mathcal{N}^{\text{rail}}$. During a day of operations, flights are operated by aircraft, and rail legs by trains. Let $\mathcal{L}$ denote the index set of legs partitioned into the flight index set, $\mathcal{L}^{\text{air}}$, and the rail-leg index set, $\mathcal{L}^{\text{rail}}$. An operational day can be represented by a time-expanded network, as illustrated for one train and one aircraft in Figure 3.2. Each train and each aircraft operate a sequence of legs from $\mathcal{L}$. Between each operated leg, the aircraft (respectively train) stops at an airport (respectively train station), to disembark arriving passengers and embark departing passengers for the next flight (respectively rail-leg). This duration is called turnaround time (TAT) for flight, and dwell time (DW) for trains. Let $\mathcal{P}^{\text{air}}$ and $\mathcal{P}^{\text{rail}}$ be the index sets of flight pairs and rail-leg pairs operated consecutively by the same vehicle. For each station $n \in \mathcal{N}$, we define $\mathcal{L}_n^A$ and $\mathcal{L}_n^D$ the index sets of the legs arriving at station n and departing from station n, respectively.

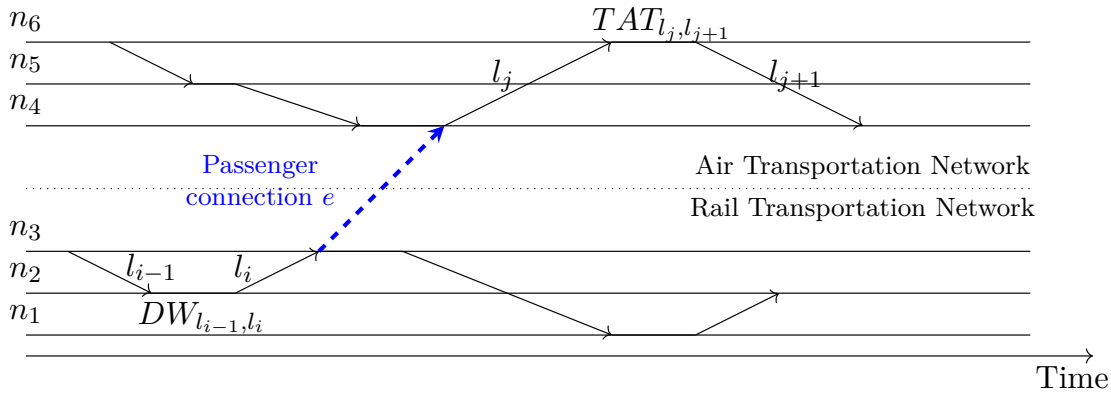


Figure 3.2 – Time-expanded flight and rail-leg networks of one train and one aircraft. The dashed line illustrates a passenger connection from a rail leg, l_i, to a flight, l_j.

3.1.2.2 Modelling passenger connections

Travellers can use one or more legs (rail or flight) to reach their final destination. If there is no direct leg, they have to connect between two legs. For instance, a passenger connection between a train and a flight is represented by the dashed line in Figure 3.2. A passenger connection corresponds to a link between two legs l_i and l_j, $(l_i, l_j) \in \mathcal{L} \times \mathcal{L}$. Let $\mathcal{E}$ denote the index set of passenger connections. A passenger connection $e \in \mathcal{E}$ is represented by a tuple $e = (l, l', d_e, \theta_e) \in \mathcal{L} \times \mathcal{L} \times \mathbb{N} \times \Theta$, where l and l' are the first leg

and second leg, respectively ; d_e is the passenger volume demand for that connection, and θ_e is the connection type. The connection type corresponds to the type of legs involved in the connection and must be taken into account when determining the optimal connection time. In fact, depending on the transportation modes used, additional station processing times must be included in the connection time. For instance, passengers connecting from a train to a flight should allow enough time for check-in, border controls and airport security process. On the other hand, passengers connecting between two flights usually complete these steps at the same airport, reducing the required connection time. Therefore, it is important to consider the type of connection in order to provide passengers with a smooth transfer between legs.

In our study, we aim at improving connections for passengers transferring between rail and air, while maintaining connectivity for air transferring passengers. Indeed, air connecting passengers at hub airports are a priority for airlines, and flight schedules are commonly designed to facilitate these transfers. We therefore propose a model that guarantees these connections for air transferring passengers, even after the rescheduling process. Thus, the index set of passenger connections, $\mathcal{E}$, is partitioned into the index sets $\mathcal{E}^{\text{rail,air}}$, which refers to multimodal passenger connections, and $\mathcal{E}^{\text{air,air}}$ for air-air passenger connections. Multimodal connections correspond to passengers transferring from a train to a flight and vice versa.

Finally, as explained above, in the case of a multimodal transfer, the train station is not necessarily located at the airport. Consequently, we consider also the transfer time between stations in the model. Let $\pi_{l,l'}$ denote the transfer time between the arrival station of leg l and the departure station of leg l', $(l, l') \in \mathcal{L} \times \mathcal{L}$. If the transfer occurs within the same station, or if the train station is located at the airport, the transfer time between the airport and the train station is assumed to be zero. Note that in practice, depending on the airport configuration, it may in fact require some times for passengers to reach their departure terminal from the train station, even if it is directly located at the airport. This parameter can easily be set by the final user to a value different from zero. For a train station that is not located at the airport, the transfer time is obtained using the public transport schedule between the airport and the train station.

3.1.2.3 Operational constraints

In order to propose a realistic model, several operational constraints are considered in this thesis.

3.1.2.3.1 Airport slot allocation A full day (consisting of 1440 minutes) is divided into h-minute intervals, where h is a divider of 1440. In the following, we shall use $h = 5$. Let $\mathcal{T} = \{0, 1, \ldots, \frac{1440}{h}\}$ denote the index set of time steps. In order to ensure safety, airports limit the number of aircraft movements (takeoffs and landings) per time intervals, which are commonly referred to as *slots*. A *slot* corresponds to a period of time at which an

aircraft is authorised to use the full airport infrastructure to depart or arrive (International Air Transport Association (IATA), 2023b). Indeed, depending on the demand, the airport capacity can be regulated for safety purposes. Airports are therefore partitioned into three categories according to the following IATA definition :

- Level 1 airport : "*A Level 1 airport is one where the capacity of the airport infrastructure is generally adequate to meet the demands of airport users at all times*" ;
- Level 2 airport : "*A Level 2 airport is one where there is potential for congestion during some periods of the day, week, or season, which can be resolved by schedule adjustments mutually agreed between the airlines and facilitator*" ;
- Level 3 airport : *"A Level 3 airport is one where :*
  - *Demand for airport infrastructure significantly exceeds the airport's capacity during the relevant period ;*
  - *Expansion of airport infrastructure to meet demand is not possible in the short term ;*
  - *Attempts to resolve the problem through voluntary schedule adjustments have failed or are ineffective ; and as a result, a process of slot allocation is required whereby it is necessary for all airlines and other aircraft operators to have a slot allocated by a coordinator in order to arrive or depart at the airport during the periods when slot allocation occurs".*

Airlines are allocated these slots months in advance, based on their requested departure and arrival times. Generally, slots correspond to 10- and 60-minute interval times (COHOR - Association pour la coordination des horaires, 2023; Fluko - Flughafenkoordination Deutschland GmbH, 2023). Consequently, for level-3 airports, the number of movements will be limited to avoid congestion. The airport movement capacities are collected on the national authorities' websites (AECFA, 2023; COHOR - Association pour la coordination des horaires, 2023; Fluko - Flughafenkoordination Deutschland GmbH, 2023). The most recent values available for each airport considered here are summarised in Tables Appendix B. If the values for 2019 and 2021 are available, these values are retained. If not, we used the data of 2023. The values for CDG, FRA and MAD airports are represented in Figure 3.3. In the following, W denotes the set of slot durations expressed in number of time steps. Note that each element $w \in W$ must be defined so that $\frac{|\mathcal{T}|-1}{w}$ is integer. In our application context, the slots considered are 10- and 60-minute interval times. Therefore, $W = \{\frac{10}{h}, \frac{60}{h}\}$, where h is the discretisation time-step. In order to ensure these slot capacity constraints at airports, for each $w \in W$, we define $\mathcal{T}_w = \{1, w+1, 2w+1, \ldots, |\mathcal{T}| - w\}$ the set of time-step indices corresponding to the beginning each slot. For instance, with a discretisation time-step $h = 5$ minutes, if we want to count the number of aircraft that are scheduled to take-off per 10-minute intervals between 1pm and 2pm, 1pm corresponds to time step 156, 2pm corresponds to time-step 168 and $W = \{2\}$ ($\frac{10}{5}$). Then, we must ensure that the slot capacity constraints are satisfied for every time slot in $T_2 = \{156, 158, 160, 162, 164, 166\}$.

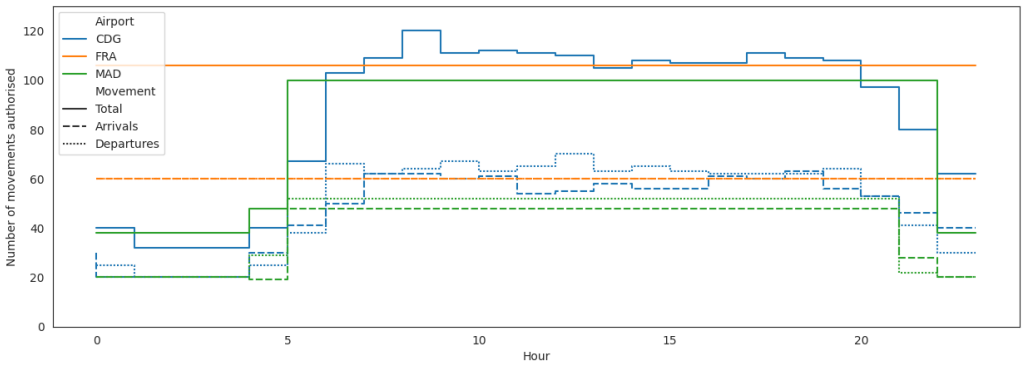


Figure 3.3 – Maximal number of movements (arrivals, departures and total) authorised per hour at the three considered hubs : CDG, FRA and MAD.

3.1.2.3.2 Track capacity at train stations Similarly, train stations have a limited capacity. For simplicity, we only assume here that the number of trains stopped at the station cannot exceed the number of platforms. We do not consider headway between trains, but they could be included by adding a minimum separation constraint between two consecutive departure times. For each train station $n \in \mathcal{N}^{\text{rail}}$, $O_n^{\max}$ denotes the maximum number of trains that can be stopped at the station at the same time. More precisely, this value corresponds to the number of tracks that can be directly accessed from a platform, at the station. At each time step, the number of trains stopped at the station must be less than or equal to this value. Note that for trains that spend the night at the train station, we assume that they are not stored on the track during the night until their scheduled departure times. In general, trains are routed to the platform a few minutes before departure. In the following, we assume that trains stored at the train station will arrive on the track r minutes before the scheduled departure time, where r is chosen so that h is a divider of r (recall that h is the discretisation time step). Similarly, for trains whose train station n is the last served of the day and which will then spend the night at n, we assume that they will leave the track r minutes after their arrival time at the station. These particular trains are included, after a pre-processing step, in the departing and arriving set of train stations, $\mathcal{L}_n^D$ and $\mathcal{L}_n^A$, $n \in \mathcal{N}^{\text{rail}}$, to account for the train station occupancy at each time step. Figure 3.4 illustrates, through an example with two trains and three train stations, the computation of the train station occupancy. In this example, the time is divided into 5-minute intervals (*i.e.*, $h = 5$), and r is set to 20 minutes. Trains therefore arrive on the track four time steps before their first departure time of the day, and leave the track four time steps after their last arrival time of the day.

3.1.2.3.3 Air connectivity constraints and limited timetable adjustments As mentioned above, the model must preserve passenger air-air connections. Hence, in addi-

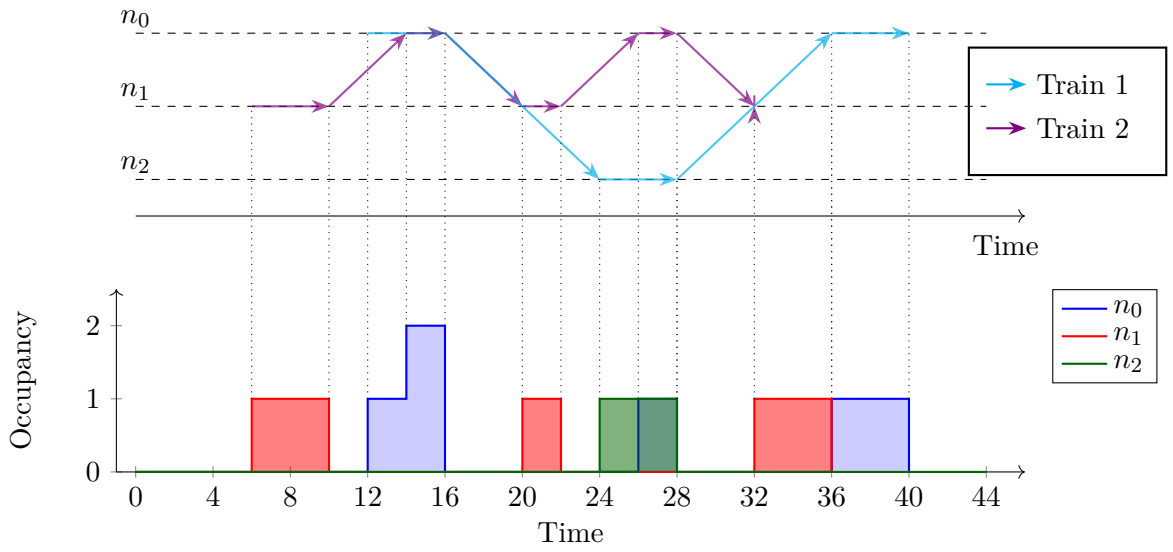


Figure 3.4 – Illustration of the train station occupancy, for three train stations, n_0, n_1 and n_2 served by two trains.

tion to guarantee that their duration is within the interval [MCT,MACT], we limit the connection time deviation, that are to be decided, to Δ minutes, where Δ is a user-defined parameter. For instance, for passengers initially having a flight-flight connection of 120 minutes, their new connection time is constrained to be within the interval $[120-\Delta, 120+\Delta]$ minutes. Finally, as presented in Chapter 1, transportation operators are generally reluctant to alter their operations in order to synchronise with other transportation modes. To palliate this issue, we set an upper bound of $\delta^{\max}$ minutes on each of the schedule deviations to be applied to flights and trains.

3.1.2.4 Optimisation problem formulation

This subsection presents the input data, the decision variables, the objective function and the constraints of the problem. In the following, this problem is referred to as the Air-Rail Strategic Timetable Synchronisation (ARSTS) problem. The notations are summarised in Table 3.1.

3.1.2.4.1 Input data For each airport $n \in \mathcal{N}^{\text{air}}$, the arrival and departure slot capacities per slot-window $w \in W$ are known (given input data). As these values depend on the hour of the day, for each $w \in W$, and each time step $i \in \mathcal{T}_w$, $Y_n^{A,w,i}$ and $Y_n^{D,w,i}$ denote the maximum number of airport arrivals and departures that could be scheduled during the time interval $[i, i+w-1]$, respectively. For each train station $n \in \mathcal{N}^{\text{rail}}$, the number of available tracks $O_n^{\max}$ is also given.

For each leg $l \in \mathcal{L}$, T_l^A denotes the initial scheduled arrival time of l. Similarly, T_l^D corresponds to the initial scheduled departure time of leg l, $l \in \mathcal{L}$. For each $l \in \mathcal{L}$, IVT_l denotes the in-vehicle time of leg l. For each pair of consecutive flights $(l, l') \in \mathcal{P}^{\text{air}}$ operated by a same aircraft, we set $MTAT_{l,l'}$ the minimum turnaround time at the airport between the two flights (the initial turnaround time planned by airlines). Similarly, for each consecutive rail-leg pair $(l, l') \in \mathcal{P}^{\text{rail}}$ operated by a same train, let $DW_{l,l'}$ denote the dwell time at the train station between leg l and l'.

3.1.2.4.2 Decision variables Decision variables of the ARSTS problem are defined as follows. For each leg $l \in \mathcal{L}$, we define the discrete decision variable, k_l^D, to be the index of time step at which l is scheduled to depart. From this primary decision variable, several auxiliary optimisation variables are defined. First, for each $l \in \mathcal{L}$, we set k_l^A the index of time step at which l is scheduled to arrive. The decision variables t_l^A and t_l^D are the new scheduled arrival time and departure time of l, $l \in \mathcal{L}$. From these variables, we further define the decision variable δ_l, the schedule deviation from the initial departure time of l, $l \in \mathcal{L}$. For each time step $i \in \mathcal{T}$, and each leg l, $l \in \mathcal{L}$, $x_{l,i}^D$ and $x_{l,i}^A$ denote binary decision variables indicating whether leg l is scheduled to depart after time step i, and scheduled to arrive after time step i, respectively. For each train station $n \in \mathcal{N}^{\text{rail}}$, and at each time step $i \in \mathcal{T}$, we define a discrete decision variable $o_{n,i}$ that counts the number of trains stopped at n at time step i. Finally, for each passenger connection $e \in \mathcal{E}^{\text{rail,air}}$, the continuous decision variable c_e denotes the passenger connection cost.

Table 3.1 – Notations of the air-rail strategic timetable synchronisation problem.

Sets and Parameters	
$\mathcal{N} = \mathcal{N}^{\text{air}} \cup \mathcal{N}^{\text{rail}}$	index set of stations partitioned into the airport index set ($\mathcal{N}^{\text{air}}$) and the train station index set ($\mathcal{N}^{\text{rail}}$)
$\mathcal{L} = \mathcal{L}^{\text{air}} \cup \mathcal{L}^{\text{rail}}$	index set of legs partitioned into the flight index set ($\mathcal{L}^{\text{air}}$) and the rail-leg index set ($\mathcal{L}^{\text{rail}}$)
$\mathcal{L}_n^A$	index set of legs scheduled to arrive at station n, $n \in \mathcal{N}$
$\mathcal{L}_n^D$	index set of legs scheduled to depart from station n, $n \in \mathcal{N}$
$\mathcal{P}^{\text{air}}$	index set of flight pairs operated consecutively by a same aircraft
$\mathcal{P}^{\text{rail}}$	index set of rail-leg pairs operated consecutively by a same train
$\mathcal{E} = \mathcal{E}^{\text{rail,air}} \cup \mathcal{E}^{\text{air,air}}$	index set of passenger connections partitioned into the set of passengers connecting from a train to a flight and vice versa ($\mathcal{E}^{\text{rail,air}}$) and the set of passengers connecting from a flight to a flight ($\mathcal{E}^{\text{air,air}}$)
h	discretisation time step (must be a divider of 1440)
$\mathcal{T}$	index set of time steps
W	set of airport slot window durations
$\mathcal{T}_w$	set of time-step indices recording the beginning of an airport slot of a duration $w, w \in W$

Δ	maximum connection time deviation authorised for air-air connections (in minutes)
$\delta^{\max}$	maximum leg schedule deviation (in minutes)
M	large-enough constant (*big M*)
r	number of minutes ahead their first departure time of the day at which trains are routed on the track / number of minutes after their last arrival time of the day at which trains are removed from the track

Input data

$Y_n^{A,w,i}$	maximum number of flights that could arrive at airport n, during a time window of width w starting at time step i, $n \in \mathcal{N}^{\text{air}}, w \in W, i \in \mathcal{T}_w$
$Y_n^{D,w,i}$	maximum number of flights that could depart from airport n, during a time window of width w starting at time step i, $n \in \mathcal{N}^{\text{air}}, w \in W, i \in \mathcal{T}_w$
$O_n^{\max}$	maximum number of trains that can stop simultaneously at train station n, $n \in \mathcal{N}^{\text{rail}}$
T_l^A	initial-schedule arrival time of leg l, $l \in \mathcal{L}$
T_l^D	initial-schedule departure time of leg l, $l \in \mathcal{L}$
$\pi_{l,l'}$	transfer time between the arrival station of leg l and the departure station of leg l', $(l, l') \in \mathcal{L} \times \mathcal{L}$
IVT_l	in-vehicle time of leg l, $l \in \mathcal{L}$
$MTAT_{l,l'}$	minimum turnaround time between flight l and l', $(l, l') \in \mathcal{P}^{\text{air}}$
$DW_{l,l'}$	dwell time at station between rail-legs l and l', $(l, l') \in \mathcal{P}^{\text{rail}}$
d_e	number of passengers involved in connection e, $e \in \mathcal{E}$
θ_e	connection type of connection e, $e \in \mathcal{E}$

Decision variables

k_l^A	integer, index of the time step at which l is scheduled to arrive, $l \in \mathcal{L}$
k_l^D	integer, index of the time step at which l is scheduled to depart, $l \in \mathcal{L}$
t_l^A	new scheduled arrival time of $l \in \mathcal{L}$
t_l^D	new scheduled departure time of $l \in \mathcal{L}$
$t_{l,l',1}, t_{l,l',2}, t_{l,l',3}$	auxiliary decision variables defined to measure the transfer disutility between legs l and l', $e = (l, l', d_e, \theta_e) \in \mathcal{E}^{\text{rail,air}}$
δ_l	schedule deviation from the initial departure time, $l \in \mathcal{L}$
$x_{l,i}^A$	binary, indicate whether leg l is scheduled to arrive after time step i, $l \in \mathcal{L}, i \in \mathcal{T}$
$x_{l,i}^D$	binary, indicate whether leg l is scheduled to depart after time step i, $l \in \mathcal{L}, i \in \mathcal{T}$

c_e	continuous, passenger connection cost, $e \in \mathcal{E}$

3.1.2.4.3 Objective function and constraints The objective is to generate a timetable between air and rail that minimises the total multimodal passengers' transferring disutility, while, as a secondary objective, keeping schedule deviation low.

The optimisation proposed is :

$$\min_{k,t,x,c,\delta} \sum_{e \in \mathcal{E}^{\text{rail,air}}} d_e c_e, \quad \sum_{l \in \mathcal{L}} \delta_l, \tag{3.2}$$

subject to :

$$k_l^A = k_l^D + \frac{IVT_l}{h} \qquad l \in \mathcal{L} \tag{3.2a}$$
$$t_l^D = h(k_l^D - 1) \qquad l \in \mathcal{L} \tag{3.2b}$$
$$t_l^A = h(k_l^A - 1) \qquad l \in \mathcal{L} \tag{3.2c}$$
$$t_l^D \leq T_l^D + \delta^{\max} \qquad l \in \mathcal{L} \tag{3.2d}$$
$$T_l^D - \delta^{\max} \leq t_l^D \qquad l \in \mathcal{L} \tag{3.2e}$$
$$t_l^D - T_l^D \leq \delta_l \qquad l \in \mathcal{L} \tag{3.2f}$$
$$T_l^D - t_l^D \leq \delta_l \qquad l \in \mathcal{L} \tag{3.2g}$$
$$MTAT_{l,l'} \leq t_{l'}^D - t_l^A \qquad (l,l') \in \mathcal{P}^{\text{air}} \tag{3.2h}$$
$$DW_{l,l'} = t_{l'}^D - t_l^A \qquad (l,l') \in \mathcal{P}^{\text{rail}} \tag{3.2i}$$
$$k_l^D \leq i + M x_{l,i}^D \qquad l \in \mathcal{L}, i \in \mathcal{T}, \tag{3.2j}$$
$$\sum_{i=0}^{|\mathcal{T}|-1} x_{l,i}^D = k_l^D \qquad l \in \mathcal{L}, \tag{3.2k}$$
$$k_l^A \leq i + M x_{l,i}^A \qquad l \in \mathcal{L}, i \in \mathcal{T}, \tag{3.2l}$$
$$\sum_{i=0}^{|\mathcal{T}|-1} x_{l,i}^A = k_l^A \qquad l \in \mathcal{L}, \tag{3.2m}$$
$$\sum_{\tau=i}^{i+w-1} \sum_{l \in \mathcal{L}_n^A} x_{l,\tau-1}^A - x_{l,\tau}^A \leq Y_n^{A,w,i} \qquad n \in \mathcal{N}^{\text{air}}, w \in W, i \in \mathcal{T}_w, \tag{3.2n}$$
$$\sum_{\tau=i}^{i+w-1} \sum_{l \in \mathcal{L}_n^D} x_{l,\tau-1}^D - x_{l,\tau}^D \leq Y_n^{D,w,i} \qquad n \in \mathcal{N}^{\text{air}}, w \in W, i \in \mathcal{T}_w, \tag{3.2o}$$
$$\sum_{l \in \mathcal{L}_n^D \cap \mathcal{L}_n^A} (x_{l,i}^D - x_{l,i}^A) \leq O_n^{\max} \qquad n \in \mathcal{N}^{\text{rail}}, i \in \mathcal{T} \backslash \{0\}, \tag{3.2p}$$
$$t_{l'}^D - t_l^A - \pi_{l,l'} = t_{l,l',1} + t_{l,l',2} + t_{l,l',3} \qquad e \in \mathcal{E}^{\text{rail,air}}, \tag{3.2q}$$

$$c_{\theta_e,1} t_{l,l',1} - c_{\theta_e,1} \underline{t_\theta} + c_{\theta_e,2} t_{l,l',3} \leq c_e \qquad e \in \mathcal{E}^{\text{rail,air}}, \qquad (3.2r)$$

$$\text{MCT}_{\theta_e} \leq t_{l,l',1} \leq \underline{t_{\theta_e}} \qquad e \in \mathcal{E}^{\text{rail,air}}, \qquad (3.2s)$$

$$0 \leq t_{l,l',2} \leq \overline{t_{\theta_e}} - \underline{t_{\theta_e}} \qquad e \in \mathcal{E}^{\text{rail,air}}, \qquad (3.2t)$$

$$0 \leq t_{l,l',3} \leq \text{MACT}_{\theta_e} - \overline{t_{\theta_e}} \qquad e \in \mathcal{E}^{\text{rail,air}}, \qquad (3.2u)$$

$$t^D_{l'} - t^A_l \leq T^D_{l'} - T^A_l + \Delta \qquad e \in \mathcal{E}^{\text{air,air}}, \qquad (3.2v)$$

$$t^D_{l'} - t^A_l \geq T^D_{l'} - T^A_l - \Delta \qquad e \in \mathcal{E}^{\text{air,air}}, \qquad (3.2w)$$

$$t^D_{l'} - t^A_l \leq \text{MACT}_{\theta_e} \qquad e \in \mathcal{E}, \qquad (3.2x)$$

$$t^D_{l'} - t^A_l \geq \text{MCT}_{\theta_e} \qquad e \in \mathcal{E}, \qquad (3.2y)$$

$$k^D_l, k^A_l \in \mathcal{T} \backslash \{0\} \qquad l \in \mathcal{L}, \qquad (3.2z)$$

$$t^A_l, t^D_l \in \{0, h, \ldots, h(|\mathcal{T}| - 2)\} \qquad l \in \mathcal{L}, \qquad (3.2aa)$$

$$\delta_l \in \{0, 1, \ldots, \delta^{\max}\} \qquad l \in \mathcal{L}, \qquad (3.2ab)$$

$$x^D_{l,i}, x^A_{l,i} \in \{0, 1\} \qquad l \in \mathcal{L}, i \in \mathcal{T}, \qquad (3.2ac)$$

$$0 \leq c_\theta \leq 1 \qquad e \in \mathcal{E}^{\text{rail,air}}, \qquad (3.2ad)$$

where M is some large-enough *big-M*, constant to be set by the user. One can show that it is sufficient to set $M = |\mathcal{T}|$. We set M to this value in our numerical experiments. The objective function is a bi-objective optimisation function. The first objective sums up the intermodal passenger costs, while the second objective measures for each leg its deviation from the initial scheduled departure time. Constraints (3.2a)-(3.2e) refer to the new schedule assignment and to the maximum schedule deviation limit. Constraints (3.2f) and (3.2g) measure the timetable deviation for each leg (classical linearisation of the objective-function term δ_l $(= |t^D_l - T^D_l|)$), to be minimised. Constraints (3.2h) and (3.2i) stipulate that the aircraft minimum turnaround time and the train constant dwell time constraints at station are satisfied, respectively. Constraints (3.2j) to (3.2p) ensure that the number of scheduled flights per time interval does not exceed the airport capacity, and that the number of trains scheduled to stop at a station does not exceed the number of tracks. For instance, equation (3.2n) sums the number of flights scheduled to arrive over the time intervals of the day. If, for instance, the number of arrival is measured per 60-minute interval, as h is then set to 5 minutes, we fix $w = 12$. Thus, the number of arrivals is counted between time steps 1 and 12, then between time steps 13 and 24 and so on. Remark that the summation counter τ is always a valid index since w is defined so that $\frac{|\mathcal{T}|-1}{w}$ is integer. Constraints (3.2o) are similarly defined for the airport departure throughput evaluation. Constraints (3.2q) to (3.2u) are classical linearisation constraints to account for the convex piecewise-linear cost function c_θ. Constraints (3.2v) and (3.2w) guarantee air-air connections for passengers, while constraints (3.2x) to (3.2y) ensure the passenger transfer feasibility. Finally, constraints (3.2z) and (3.2ad) specify the definition domain of the decision variables. Remark that, in the formulation, for each leg $l \in \mathcal{L}$, we do not constrain decision variables k^D_l and k^A_l

to be integer since constraints (3.2j)-(3.2m) ensure that k_l^D and k_l^A will automatically be integers (sum of binary decision variables).

3.1.2.5 Bi-criterion resolution approach

As a reminder, the problem involves a bi-criterion objective function. The first criterion is the passenger disutility, defined in terms of multimodal transfer times, while the second criterion is the total changes in the schedule of airlines and railway operators (in minutes), both to be minimised. As it is common in practical multicriterion optimisation, these two objectives are competing. Here, we prioritise the passenger criterion, and the schedule deviation is considered as a secondary objective. In order to solve such *lexicographic-order* optimisation problems, two strategies are generally proposed in the literature (Chankong and Haimes, 1983). The first method consists in *scalarising* the problem, through the consideration of a weighted sum of the criteria. The second strategy is referred to as the ϵ-*constraint method*. It involves successively minimising one criterion, with the additional constraint that the value of the other criterion is not greater than ϵ. The advantage of such a sequential optimisation process is that no weighting parameter has to be managed by the user, facilitating its application. This is the method chosen here. Let $\mathcal{P}_0$ and $\mathcal{P}_1$ denote the mono-criterion problems of minimising criteria f_0 and f_1, respectively (with all the constraints seen before). In the following, let $f_0^{\mathcal{P}} = \sum_{e \in \mathcal{E}^{\text{rail,air}}} d_e c_e$ denote the optimal value of the total passenger disutility for problem $\mathcal{P}$. Similarly, we define $f_1^{\mathcal{P}} = \sum_{l \in \mathcal{L}} \delta_l$, the optimal schedule deviation of the solution for $\mathcal{P}$. As explained earlier, the passenger criterion is prioritised. Hence, we propose solving $\mathcal{P}_0$ first, and the solution value $f_0^{\mathcal{P}_0}$ is obtained. Then, $\mathcal{P}_1$ is solved with the additional constraint that the passenger criterion value is not degraded. In our test we shall add constraint $f_0^{\mathcal{P}_1} \leq f_0^{\mathcal{P}_0} + 10^{-2}$ to $\mathcal{P}_1$.

Numerical results are presented in the following section.

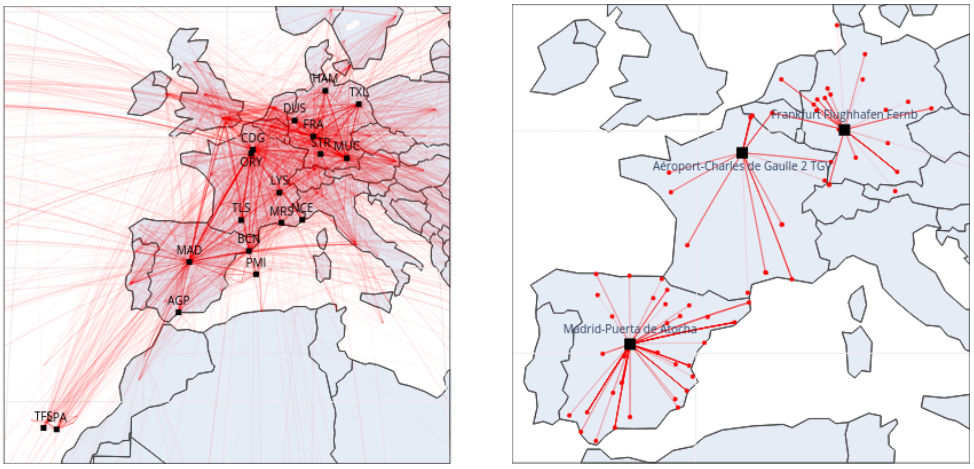
3.1.3 Application to the Western Europe transportation network

In this section, the synchronisation methodology presented above is tested on the case study of Western Europe. The framework of the study is first detailed, then the ARSTS problem is solved on passenger instances generated in Chapter 2. Numerical results and analyses of the obtained solutions are presented in Subsections 3.1.3.3 to 3.1.3.6. All the computations are performed on a laptop equipped with an AMD Ryzen 5 4500U CPU and 16 GB RAM.

3.1.3.1 Case study presentation

The multimodal network studied is composed of 18 airports : the six largest airports of each of the three countries France, Germany and Spain. We consider for each day between December 2^{nd}, 2019 and December 7^{th}, 2019, and on September 22^{nd}, 2021, all flights

operated by an aircraft that stops in one of these 18 airports. Eurocontrol historical flight schedules are used for the study (Eurocontrol, 2023). On average, 10,153 flights per day are considered in 2019, compared with 6,783 on September 22^{nd}, 2021. A particular focus is made on three airports : Paris-Charles de Gaulle airport (CDG), Frankfurt airport (FRA) and Madrid-Barajas airport (MAD). These are hub airports with several thousands of connecting passengers each day. In addition, CDG and FRA have a direct access to an HSR station : Aéroport-Charles de Gaulle 2 TGV station and Frankfurt Flughhafen Fernbahnhof. Hence, passengers arriving from a train can directly connect with a flight. On the contrary, MAD airport is not equipped with a direct HSR station. In the following, we consider connections with trains arriving at Madrid-Atocha train station. For each of these three above-mentioned train stations, trains scheduled to stop during the day are included in the study. For passengers arriving at Madrid-Atocha, an additional transfer time of 45 minutes to reach the airport from the train station is added (RENFE, 2022). Train schedules are obtained from GTFS data, as detailed in Chapter 2. This leads to consider 561 rail legs per day on average in 2019, and 473 in 2021. The air transportation network and the rail transportation network of December 2019 are presented in Figure 3.5.



(a) Air transportation network **(b)** Rail transportation network

Figure 3.5 – Transportation networks considered for December 2019. Airports and train stations are represented by black points, rail legs and flights by red lines. The width of the red lines increases with the frequency of legs operated between each OD pair per day.

The timetable synchronisation problem is then tested to coordinate flights at these three hubs with their associated train stations. Note that airport *capacity* constraints are only evaluated on the 18 airports. Similarly, train station *capacity* constraints are evaluated on the three train stations presented in the case study. However, aircraft turnaround times and train stop duration constraints are taken into account at every airport and train station

visited by an aircraft or a train. As an example, on December 2^{nd}, 2019, this involves a total of 496 airports and 72 train stations. If we compare the maximum hourly capacity declared for each airport with the actual number of movements, the actual volume is generally lower than declared. However, it may happen that the hourly volume of operated flights was higher than the declared one. In order to obtain a feasible solution, the hourly capacity limits are therefore set to the maximum between the number of movements historically observed and the capacity declared for airports.

In this section, the ARSTS problem is solved for each of the 60 generated passenger demand instances of 2019, and the 10 instances generated for 2021. In Europe, there is an agreement between 26 countries to abolish border controls. These countries constitute the Schengen area, reducing thereby the airport processing time for flights operated within these countries. Such flights are referred to as Schengen flights, and flights with a destination outside the Schengen area are referred to as non-Schengen flights. Hence, we define $\Theta = \{\text{TSF}, \text{TNSF}, \text{FT}, \text{FF}\}$ the connection type sets considered, where "T-SF" correspond to train-flight connections with a flight destination within the Schengen area, "T-NSF" to train-flight connections with a flight destination outside the Schengen area, "FT" to flight-train connections and "F-F" to air connecting passengers. For each day considered in the study, the number of passengers making each connection type is given in Table 3.2.

Day	TSF	TNSF	FT	FF	Total
02-12-2019	13,388	18,302	32,000	89,915	153,605
03-12-2019	14,660	16,720	31,269	88,782	151,431
04-12-2019	13,796	17,955	31,774	89,775	153,300
05-12-2019	10,949	20,059	31,148	88,850	151,006
06-12-2019	13,609	18,115	31,692	90,205	153,621
07-12-2019	12,910	19,071	31,928	90,416	154,325
22-09-2021	6,716	4,890	11,487	31,646	54,739

Table 3.2 – Volume of connecting passengers per day.

The values of each parameter defined in Section 3.1.1 are presented in Table 3.3. For each connection type $\theta \in \Theta$, we make the assumption that the passenger disutility function is normalised to the value 1 when the transfer time is equal to either the lower-bound value MCT_θ, or to the upper-bound value MACT_θ, (see Figure 3.1). Values of MCT_θ, MACT_θ, $t_{\min}$ and $t_{\max}$ are chosen according to several studies (Dennis, 1994; Burghouwt and de Wit, 2005; Danesi, 2006), but they can be tuned by the final user depending on the airport considered.

In addition to the initial train and flight schedules, three scenarios are tested. The first scenario, S1, assumes that railway operators adapt their schedules to fixed, given flight schedules, only changes to the train schedules are authorised in the optimisation process. Similarly, scenario S2 corresponds to the case of airlines adapting their schedules

Table 3.3 – Parameters values (input data) used in the instances ; r, $\delta^{\max}$, Δ, MCT_θ, $\underline{t_\theta}$, $\overline{t_\theta}$ and MACT_θ are in minutes.

Passenger cost-function parameters							$\delta^{\max}$	Δ	r
θ	MCT_θ	$\underline{t_\theta}$	$\overline{t_\theta}$	MACT_θ	$c_{\theta,1}$	$c_{\theta,2}$			
T-SF	45	80	100	270					
T-NSF	60	110	130	300	$\frac{1}{\text{MCT}_\theta - \underline{t_\theta}}$	$\frac{1}{\text{MACT}_\theta - \overline{t_\theta}}$	30	15	20
FT	30	50	70	180					

to fixed, given train schedules, only changes in the flight schedules are authorised. Finally, scenario S3 allows one to apply changes in both train and flight schedules. This last scenario corresponds to the existence of some bilateral agreement between rail and air operators to improve the passenger connection quality.

3.1.3.2 Numerical results

For each of the subproblems $\mathcal{P}_0$ and $\mathcal{P}_1$, the average number of decision variables (and among which the number of binary variables), and constraints per instances, for either 2019 or 2021, and per scenario, are listed in Table 3.4.

Table 3.4 – Average number of decision variables and constraints for the 60 instances considered in December 2019, and the 10 instances in September 2021.

Year	**# variables (# binary)**	**# constraints**
2019	293,473 (245,692)	433,065
2021	230,375 (196,814)	317,076

Our optimisation problem is solved using the MIP solver Gurobi, version 9.1.2 (Gurobi Optimization, LLC, 2023). A time limit of 30 minutes is set for the resolution of each monocriterion subproblem P_0 (passenger disutility) and P_1 (schedule deviation). Computation times for each scenario are presented in Table 3.5. For the instances of 2019, the total computation time is around 2 seconds, 2 minutes and 6 minutes on average for scenarios S1, S2 and S3, respectively. In 2021, airport and train station capacities were largely greater than the actual number of flights and trains operated. In addition, the number of connections has decreased by over two third, compared with 2019. The computation time is therefore lower, with a solution found within 1 minute and 2 seconds on average for scenario S3. Recall that the tool is assumed to be run several weeks in advance, the computation time is thereby acceptable for practitioners. Detailed results obtained on the instances of 2019, for each individual scenario, are presented and discussed in the next two

Table 3.5 – Average computation time, μ, in seconds, and the associated standard deviation, σ, for the 60 instances considered in December 2019, for the two monocriterion subproblems P_0 and P_1.

		P_0 **(passenger disutility)**			P_1 **(schedule deviation)**			**Total**
Year	**Scenario**	μ (s)	σ (s)	MIPGap (%)	μ (s)	σ (s)	MIPGap (%)	μ (s)
2019	S1	0.7	0.1	0.0	0.9	0.2	0.0	1.6
	S2	23.4	10.4	0.4	79.5	26.1	0.8	102.9
	S3	103.4	67.4	0.8	255.5	167.8	0.8	358.9
2021	S1	0.8	0.1	0.0	1.0	0.2	0.4	1.8
	S2	12.5	1.9	0.0	25.2	13.3	0.6	25.8
	S3	26.8	19.9	0.3	60.4	35.2	0.4	62.0

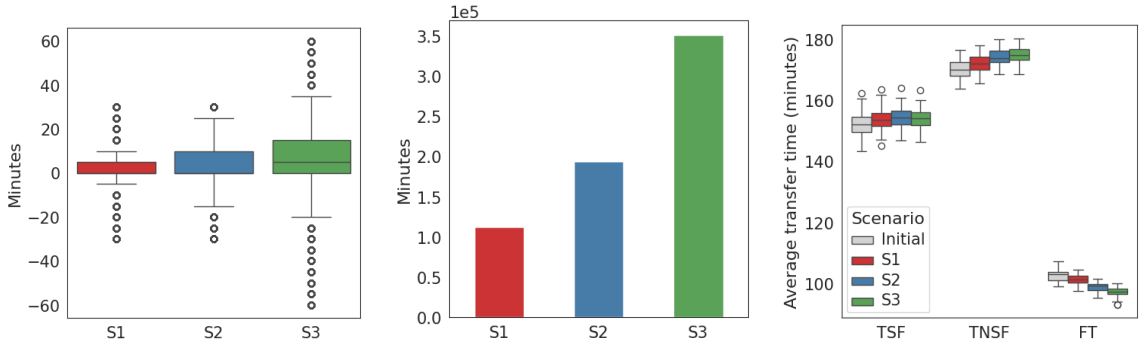
subsections. Then, a comparative analysis between the results of 2019 and 2021 is proposed in Subsection 3.1.3.5. Finally, a sensitivity analysis of the maximum leg schedule deviation parameter $\delta^{\max}$, is presented in the last subsection.

3.1.3.3 Benefits of air-rail collaboration

Figure 3.6a presents the average number of minutes gained per connection in scenarios S1, S2 and S3. For passengers initially having a short connection time ($t < \underline{t_\theta}$), these minutes correspond to additional minutes in the new schedule. Conversely, for passengers with initially a long connection time ($t > \overline{t_\theta}$), these minutes correspond to a reduction in their connection time. More precisely, for each passenger connection $e \in \mathcal{E}^{\text{rail,air}}$, let t_e^{init} denote the initial duration of the connection, and t_e^{optim}, the connection duration with the new schedule. Let g_e denote the number of minutes gained for connection $e, e \in \mathcal{E}^{\text{rail,air}}$; it is computed as follows :

$$g_e = \begin{cases} t_e^{\text{optim}} - t_e^{\text{init}}, & \text{if } t_e^{\text{init}} \in [\text{MCT}_{\theta e}, \underline{t_{\theta_e}}] \\ t_e^{\text{init}} - t_e^{\text{optim}}, & \text{if } t_e^{\text{init}} \in [\underline{t_{\theta_e}}, \overline{\text{MACT}_{\theta e}}]. \end{cases} \tag{3.3}$$

On average, 5.6 minutes per connection are thereby gained when both airlines and railway operators cooperate to change their schedules. If only the train schedule is modified, 1.8 minutes could be gained on average per connection, compared with 3.0 minutes if only the flight schedule is changed. As expected, changing both rail and air schedules allows one to obtain a higher benefit for passengers, since this improves (reduces or increases) the connection time by up to one hour, compared with only 30 minutes if we authorise changes in only one mode. Moreover, 50% of transferring passengers gain more than 5 minutes per connection in scenario S3. Figure 3.6b displays the total number of minutes gained per day, per scenario. Scenarios S1 and S2 already improve the passenger experience by



(a) Total number of minutes gained per connection, per simulation.

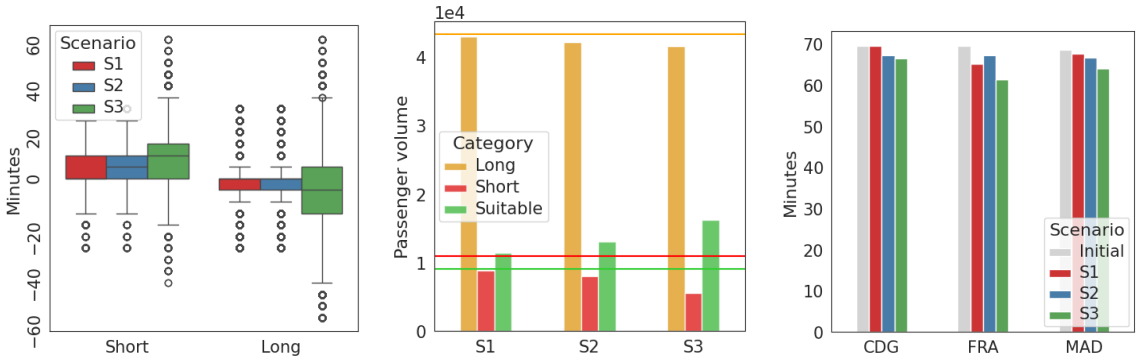
(b) Total (in 10^5 minutes) minutes gained per day, on average.

(c) Average connection time per simulation.

Figure 3.6 – Distribution of minutes gained per connection in each scenario (a), total passenger minutes gained (b) on average, and average connection time (c) for scenario S1, S2 and S3.

gaining around 1,800 and 3,200 hours for passengers, respectively. These hours correspond to initial extra wait time or stressful tight connections for passengers. However, if both airlines and railway operators collaborate, it is more than 5,800 hours that could be thereby gained for passengers. This value is even larger than the sum obtained with the two unilateral scenarios. This shows that collaboration between modes brings a substantial added value for passengers. For each considered scenario, the average connection time per simulation is displayed in Figure 3.6c. A distinction is made between connections from a train to a Schengen-destination flight (TSF), from a train to a non-Schengen-destination flight (TNSF), and from a flight to a train (FT). In the initial planning, the average connection times in the three considered hubs, for TSF, TNSF and FT connections are 152 minutes, 170 minutes, and 102 minutes, respectively. For TSF and TNSF connections, all scenarios increase the average transfer time. Indeed, as the time-unit cost ($c_{\theta,1}$) of short connections is set higher than the time-unit cost of long connections ($c_{\theta,2}$), short connections are lengthened in priority, increasing thereby the average connection time. Regarding FT connections, the average connection time is decreased in all scenarios. Table 3.2 shows that a larger number of passengers have this type of connections. Consequently, in addition to lengthen short connection times for FT connecting passengers, initially long connection times are also reduced.

Figure 3.7a displays the changes in minutes per connection, for each scenario, based on the initial category of the connection : *short* ($t < t_{\min}$) or *long* ($t > t_{\max}$). One observes that for scenarios S1 and S2, short connections are lengthened, but limited changes are applied to long connections, at the benefit of passengers having short connections. In addition, scenario S2 reduces the number of short connections more than S1 does (the median value



(a) Shift in minutes per connection.

(b) Passenger volume. Horizontal lines correspond to the initial planning.

(c) Average passenger wait time.

Figure 3.7 – Shift in connection time (a) as a function of the initial connection category (*short* and *long*), passenger volume per connection category for each scenario (b), and average wait time for passengers initially experiencing *long* connection time at the three hub airports (c).

is higher), since the number of flights is significantly higher than the number of trains, increasing thereby the magnitude of the change. Regarding scenario S3, in which both flight and train schedules are optimised, short connections are lengthened, while long connections are shortened. Figure 3.7b displays the average volume of passenger experiencing *short*, *suitable* and *long* connections, for each tested scenario. Scenario S3 accommodates more than 25% of intermodal passengers with a suitable connection time, compared with 14% in the initial planning. In the initial planning, most connections are generally long ones (more than 68%). This value is reduced to 65% after optimisation. Regarding the share of short connections, they represented more than 17% in the initial planning, on average. After optimisation, in S3, this value is reduced to 8.7%. Figure 3.7c presents the average wait time in minutes for passengers with long connections at the three considered hubs. More precisely, these minutes correspond to extra minutes (above the suitable connection time, $t_{\max}$). On average, intermodal passengers wait more than one hour for their second leg. Scenario S3 reduces the average wait time of passengers by 3, 8 and 5 minutes at CDG, FRA and MAD, respectively. These values are obtained considering passengers having initially long connection times. Figures 3.7b and 3.7c show that, even if long connections are not prioritised by the algorithm (due to the lower time-unit cost compared with short ones), both the number of long connections and their duration are reduced, by 4% and 5.2 minutes, on average.

Figure 3.8 displays the volume of passengers experiencing *short*, *suitable* and *long* connections across the day. At each hour, the share of *suitable* connections (green) has

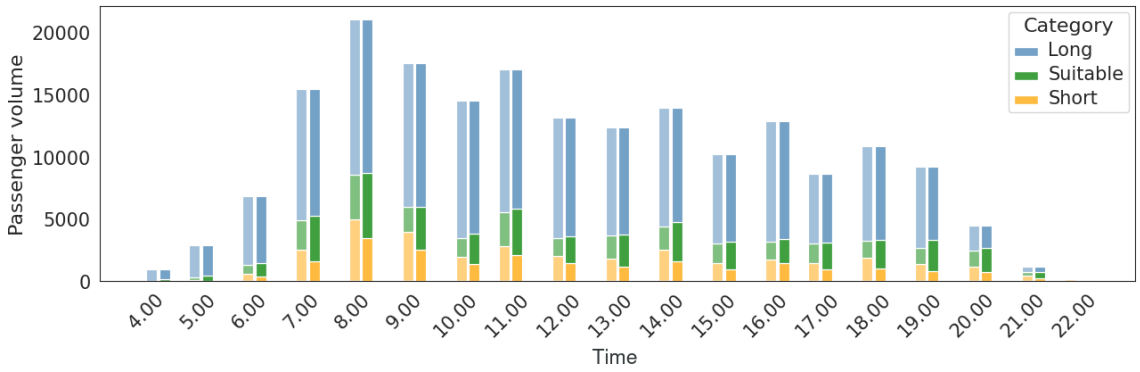
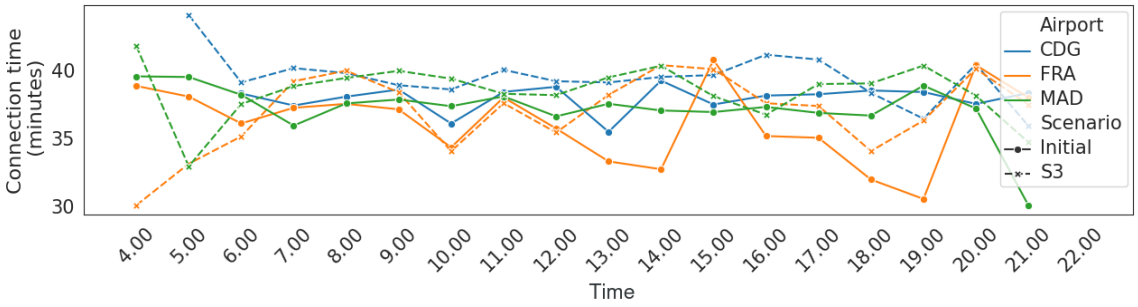


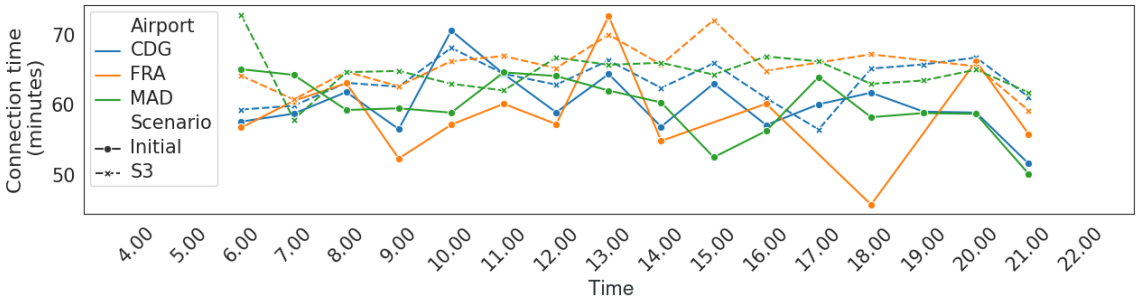
Figure 3.8 – Distribution of connection time per category (*short*, *suitable* or *long*) across the day, for each hour of the day. Transparent bars (on the left) correspond to the initial planning, opaque bars (on the right), to the optimised planning with scenario S3.

been increased. In addition the average connection time per hour, for passengers having initially short and long connection times, are displayed in Figures 3.9 and 3.10, respectively. Regarding short connections (Figure 3.9), as desired the average connection time is generally increased at each airport, and each hour, for each connection type considered. There are however a few exceptions for FRA airport : at 4am and 5am, where the average connection time of short connections was reduced on average. However, 56 passengers were initially experiencing short FT connections for flights arriving between 4am and 5am at FRA airport, and in the integrated planning, this number is reduced to 23. Similarly, 295 passengers had a short TNSF connections in the initial planning, against 13 in the S3 scenario. Regarding initially long connections, FT connections are shortened at each considered hub. However, the average connection time of long connections is stable for TSF and TNSF connections. Nevertheless, as shown by Figure 3.8, the number of long connections generally decreased at each hour.

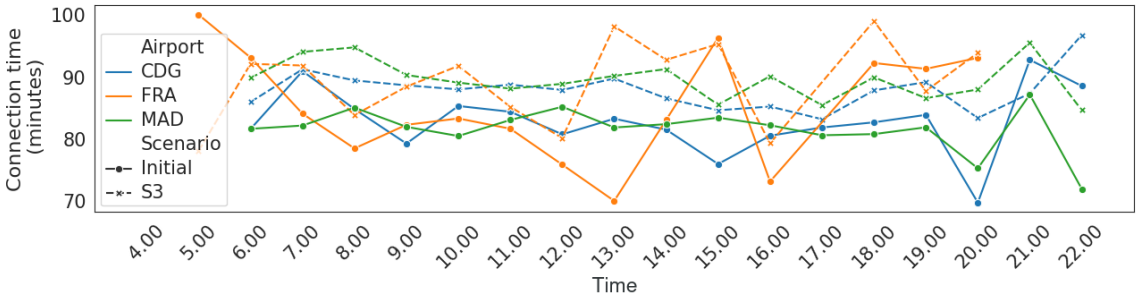
Figures 3.8, 3.9 and 3.10 highlight two main points. First, the number of *short* connections is significantly reduced in the collaborative scenario (S3), limiting the risk of missed connections for passengers. Second, for remaining *short* and *long* connections, connection times are lengthened and shortened, respectively, as desired.



(a) FT

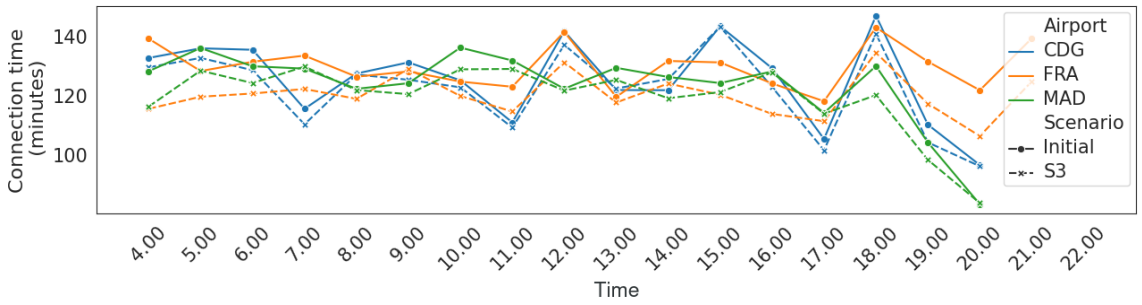


(b) TSF

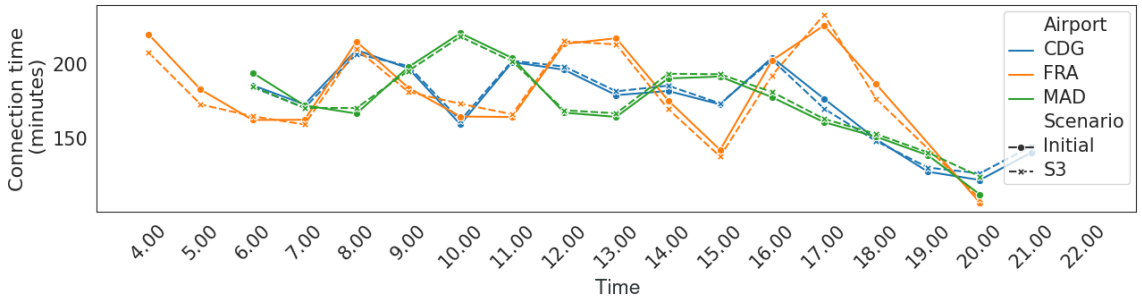


(c) TNSF

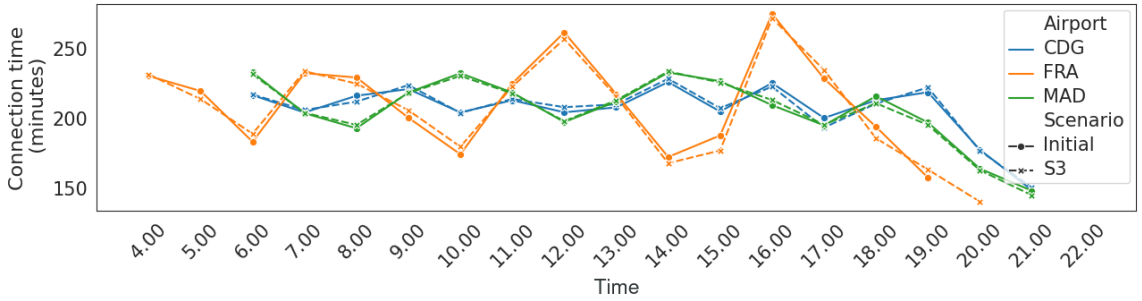
Figure 3.9 – Average connection time for passenger initially experiencing *short* connections, for FT (a), TSF (b) and TNSF (c) connection types, per hour, for the 6 days of 2019 considered.



(a) FT



(b) TSF



(c) TNSF

Figure 3.10 – Average connection time for passenger initially experiencing *long* connections, for FT (a), TSF (b) and TNSF (c) connection types, per hour, for the 6 days of 2019 considered.

Figure 3.11 presents the connection time distribution for passengers, for each connection type considered, at each hub airport, in 2019, before and after optimisation of the S3 scenario.

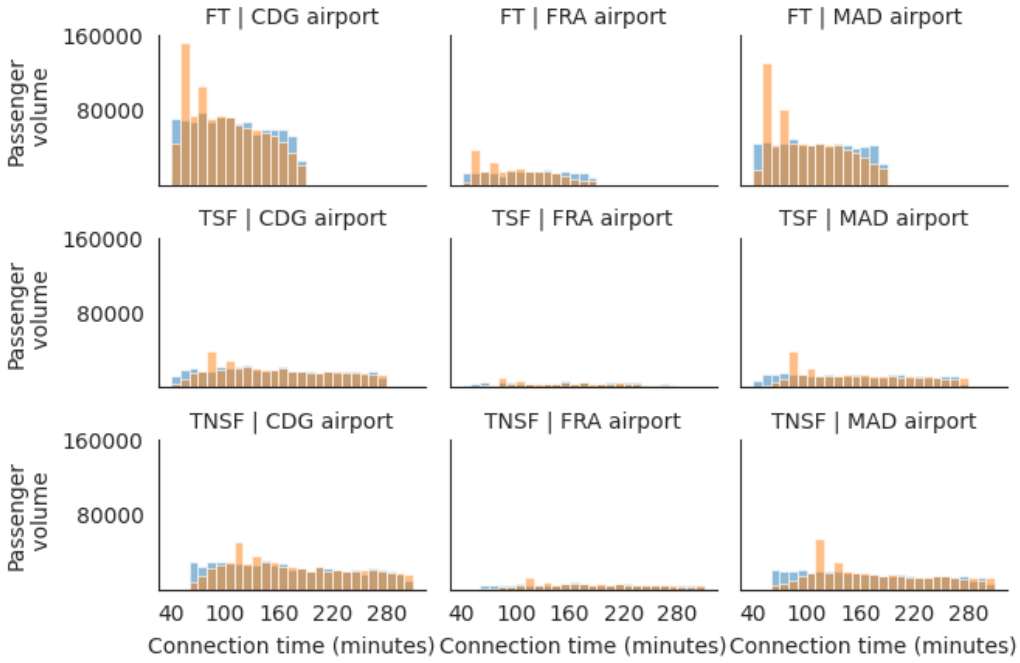


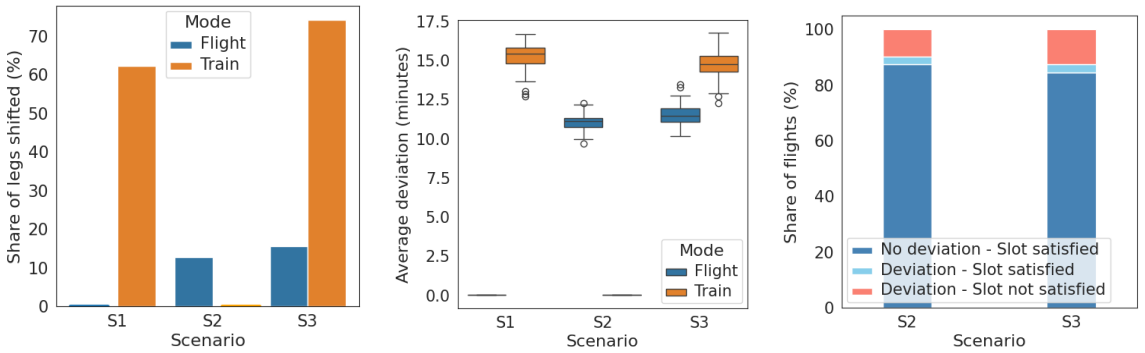
Figure 3.11 – Connection time distribution, in minutes, before (blue) and after (orange) optimisation of the S3 scenario ((brown corresponds to the superposition of blue and orange).

This figure confirms the results previously observed. Short connections are removed in priority, and for FT connections, long connection times are also reduced at the three hubs, decreasing the average connection time observed in Figure 3.6c. These FT connections are prioritised by the algorithm since, in the simulated instances, the number of passengers making FT connections is larger.

The following section presents the impact of optimising air-rail timetables on transportation schedules.

3.1.3.4 Operator costs

The average share of legs shifted, over all 60 instances, for each scenario S1, S2 and S3, is presented in Figure 3.12a. A distinction is made between trains and flights. One first observes that while more than 62% of trains are deviated from their initial schedules in the unilateral scenario S1, and 74% in the bilateral scenario S3, only 12.6% of the flights are



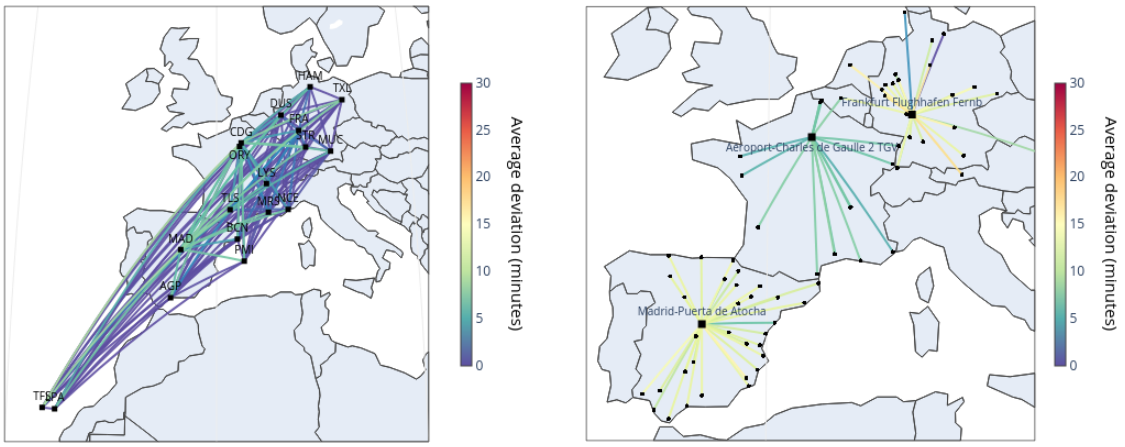
(a) Average share of legs (trains and flights) shifted.

(b) Average deviation (in minutes) per leg.

(c) Aircraft volume deviation and initial slot satisfaction.

Figure 3.12 – Deviation from initial schedule analysis for all 60 instances of December 2019.

impacted by a change in the unilateral scenario S2 and 15.4% in the bilateral scenario S3. In fact, only the trains stopping at the three train stations considered are counted, while all aircraft stopping at the 18 considered airports are taken into account. Thus, in reality, the proportion of trains shifted is smaller. A second observation is that a higher number of trains are deviated from their initial schedules in scenario S3 when compared with S1. One could have expected that allowing changes in both flight and train schedules would reduce the number of legs impacted by a schedule deviation. One hypothesis to explain this is that the change in flight schedule opens a wider range of possibility (the state space is larger). Hence, in scenario S1, changes in train schedule are determined by the initial flight schedule and airport congestion. Wide changes in the train schedule are not needed since connection quality improvement is limited, and similarly for flights in scenario S2. However, in scenario S3, the flight schedule can be changed as well. Suitable connection time can be reached, maybe at the cost of moving more trains than in scenario S1 and more flights than in S2. The average deviation (in minutes) per leg for each scenario is presented in Figure 3.12b. Note that the average deviation is computed only among deviated legs. In S2 and S3 scenarios, flights are shifted by 11.0 and 11.5 minutes from their initial schedule, on average. This value falls to 1.4 and 1.8 minutes if all aircraft (deviated or not) are considered. Regarding trains, they are modified from 15.3 to 14.7 minutes between scenarios S1 and S3, on average. Hence, although the number of deviated legs is increased, on average the range of the deviation decreases when both air and rail operators agree to change their schedules. Finally, Figure 3.12c displays the number of aircraft that are still assigned to their initial 10-minute airport slots. In scenario S2, 21.1% of deviated flights still satisfy their initial slot departure time, against 19.5% in scenario S3. Results presented in Buire *et al.* (2024) show that the average flight deviation was reduced in



(a) Flights **(b)** Trains

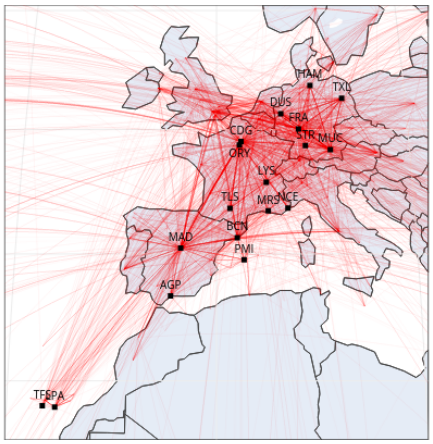
Figure 3.13 – Average train and flight schedule deviations per OD pair, for scenario S3, over the 60 instances.

S3. However, the airport capacity constraint was based the initial flight schedule. After a further analysis, it has been shown that they under-evaluate airport capacities. Hence, by setting the airport slot capacities to the ones that have been declared (Appendix B), we obtain a larger deviation of the flight schedule in S3 than in S2, for the reason cited above.

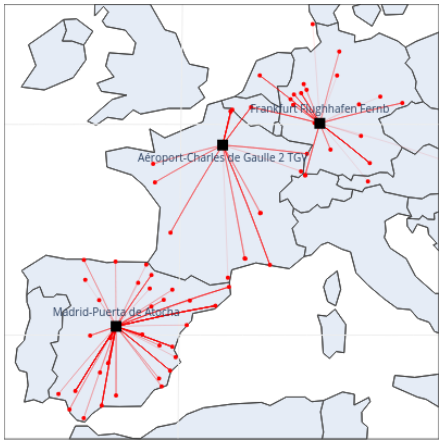
Figure 3.13 displays the average deviation over the 60 instances per OD pairs for both flights and trains for scenario S3. For the sake of clarity, only flights between the 18 airports considered are displayed. Within the considered airport network, changes in schedule are, on average, below 15 minutes. The most impacted OD flight pairs are CDG - Gran Canaria Tenerife Sur airports and MAD - Gran Canaria Tenerife Sur airport airports. Regarding train schedules, the less coordinated legs are MAD - Murcia for Spain, FRA - Innsbruck Hbf for Germany and CDG - Montpellier in France.

3.1.3.5 Comparison between 2019 and 2021

The previous subsections highlighted the potential gain of schedule synchronisation for passengers in terms of connection comfort, in 2019. This subsection provides a comparative analysis with the results obtained for 2021. As detailed above, the COVID-19 crisis deeply changed transportation network structures. Train and flight frequencies were reduced, as depicted in Figure 3.14. The flight density has decreased, with an average volume of 10,153 flights per day in December 2019, to 6783 for September 22, 2021. Recall that only flights operated by an aircraft stopping at least once at one of the 18 considered airports, during the day, are considered. Table 3.2 and results of Chapter 2, also show a significant decrease in the volume of intermodal connecting passengers, in particular in Madrid.

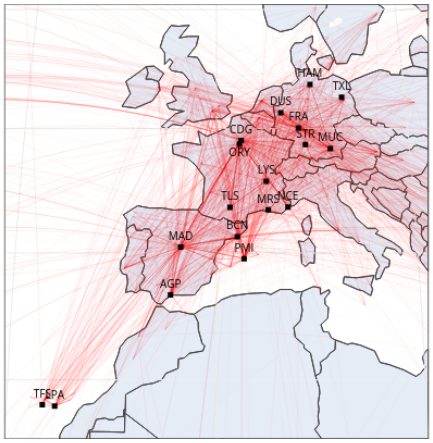


(a1) Air transportation network

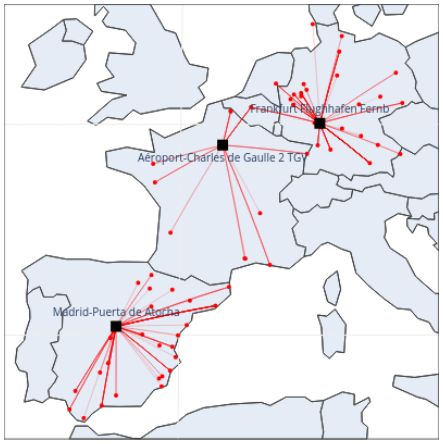


(a2) Rail transportation network

(a) December 2019



(b1) Air transportation network



(b2) Rail transportation network

(b) September 2021

Figure 3.14 – Transportation networks considered for December 2, 2019 (top) and for September 22, 2021 (bottom). Airports and train stations are represented by black points, rail legs and flights by red lines. The width increases with the frequency of legs operated between each OD pair per day.

The connection time distributions, for the 10 instances of 2021 tested, before and after optimisation, are presented in Figure 3.15. On average, in scenario S3, 5.55 minutes are saved for passengers in 2019, against 9.1 minutes in 2021. As airports and train stations are less congested, there are more opportunities for rescheduling and improving the passenger

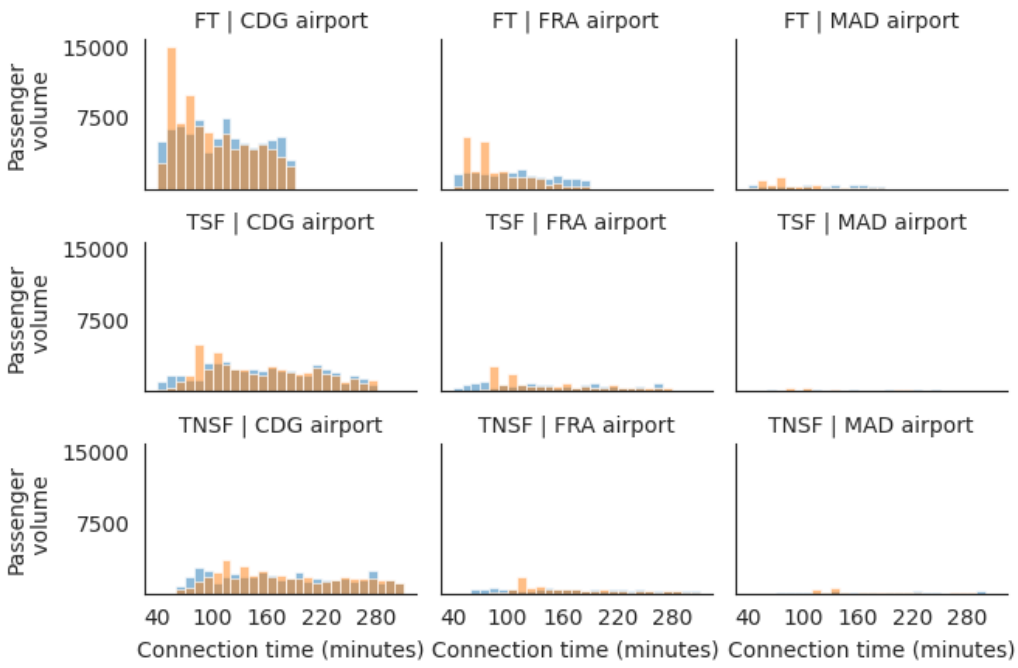


Figure 3.15 – Connection time distribution, in minutes, before (blue) and after (orange) optimisation for the S3 scenario (brown corresponds to the superposition of blue and orange).

experience. At the same time, since the number of passenger connections was reduced in 2021, the share of passenger connections that might be competing decreased. A pairwise competing connection corresponds to a connection pair that shares a common leg, where rescheduling the common leg increases the quality of one link and decreases the quality of the other. For instance, suppose a first train arrives at 10am and a second train arrives at 12am, and both trains have connecting passengers with a flight departing at 13am. If the flight is rescheduled earlier, the connection for passengers of the second train will be shortened, while being already short. On the other hand, if the flight is postponed after 13am, passengers arriving from the 10am train will have an even longer connection. Figure 3.16 presents the number of pairwise-competing connections for each leg on the (simulated) passenger demand on December 2, 2019 and September 22, 2021. In 2021, as the volume of connecting passengers decreases, the number of pairwise-competing connections also decreases. Consequently, the optimisation succeeds in saving more minutes on average per passenger in 2021 than in 2019. This results in a higher share of suitable connections in the optimised planning in 2021, compared with 2019. Table 3.6 summarises the share of each connection category (short, suitable and long) for the initial planning and for the optimal planning obtained for the S3 scenario, for 2019 and 2021. On average, for

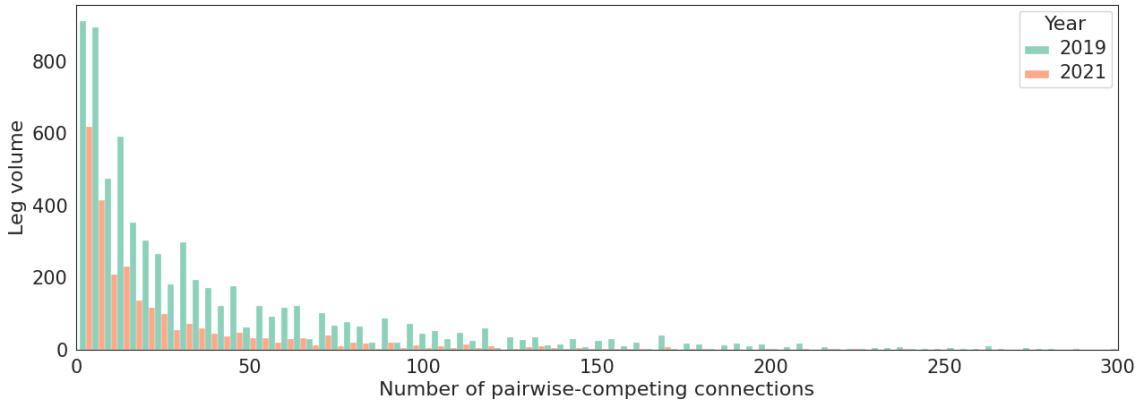


Figure 3.16 – Distribution of pairwise-competing connections per leg, for December 2, 2019 and September 22, 2021. Only values above 0 are displayed.

each instance, the share of suitable connections after optimisation is of 29.7% in 2021 and 14.1% in the initial planning, compared with 25.6% in 2019 after optimisation and 14.3% before. The increase in suitable connections is therefore higher for 2021. Regarding short connections, their proportion was lower in 2021, at the benefit of long ones. Indeed, the initial number of suitable connections is stable between 2019 and 2021, but the share of short ones decreases while the share of long ones increases. In 2021, the optimisation process succeeds in reducing by 56% the number of short connections for passengers, compared with 49% in 2019. Similarly, the number of long connections was reduced by 10% in 2021 and by 4% in 2019.

	2019		**2021**	
	Initial	S3	Initial	S3
Long	68.6	65.7	70.6	63.6
Short	17.2	8.7	15.3	6.7
Suitable	14.3	25.6	14.1	29.7

Table 3.6 – Share (%) of *suitable*, *long* and *short* connections in the initial planning and the S3-scenario optimal solution, for 2019 and 2021.

Regarding the operator cost, the share of trains and flights deviated is lower compared with 2019, as it can be observed in Table 3.7. However, the average deviation is higher. Indeed, as explained above, the over-capacited stations allow one to shift trains and flights more easily than in 2019.

	2019	**2021**
Average share of trains deviated (%)	74	56.1
Average share of flights deviated (%)	15.4	10.3
Train average deviation (minutes)	14.7	20.9
Flight average deviation (minutes)	11.5	12.9

Table 3.7 – Comparative analysis between 2019 and 2021, for scenario S3. The average deviation is computed among the legs deviated from their initial scheduled departure time.

3.1.3.6 Sensitivity analysis of the maximum schedule deviation parameter $\delta^{\max}$

This subsection analysis the sensitivity of the synchronisation tool with respect to the value of the parameter $\delta^{\max}$. For the day of December 2^{nd}, 2019, our optimisation methodology is tested on the 10 passenger demand instances, by setting $\delta^{\max}$ to the four possible values : 0 (initial planning), 30 (proposed default value), 45, and 60 minutes.

Figure 3.17 displays the value of the two optimisation criteria (passenger disutility and schedule deviation) as a function of $\delta^{\max}$, for each instance. One observes that as the

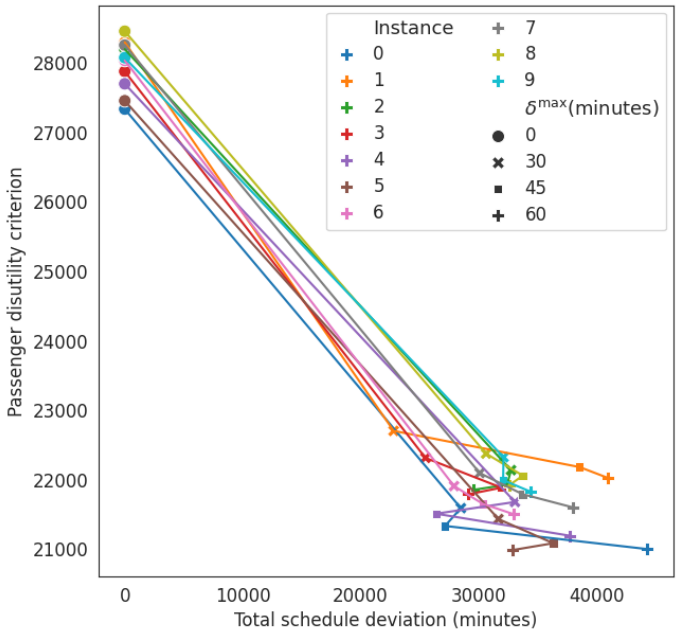


Figure 3.17 – Passenger transfer disutility criterion and schedule deviation criterion as a function of the $\delta^{\max}$ parameter, for the 10 passenger demand instances of December 2, 2019.

value of $\delta^{\max}$ increases, the passenger transferring disutility value decreases. Conversely, as expected, the value of the deviation criterion increases, since larger changes are now authorised. However, for instances 0, 2 and 4, setting the value of $\delta^{\max}$ to 45 minutes succeeds in improving both criteria, compared with $\delta^{\max} = 30$. This observation remains true for instance 2 when increasing $\delta^{\max}$ further to 60 minutes : both criteria values are lower than when $\delta^{\max}$ is set to 30 and 45 minutes, although one could have expected that increasing the magnitude of change will increase the total schedule deviation.

Table 3.8 presents the average leg deviation, in minutes, and the share of legs deviated from their initial schedule when $\delta^{\max}$ is set to 30 and 45 minutes for a typical instance : instance 1. If the parameter is set to 45 minutes, there are fewer legs deviated, compared with $\delta^{\max} = 30$, but with a larger average deviation. This shows that authorising a larger schedule deviation (larger value of $\delta^{\max}$) may allow an even larger improvement for some passengers, at a global lower cost for the operators. Nevertheless, this solution is less fair among legs, since although a lower number of legs are impacted by a change, their magnitude of change is higher. The choice of the value of $\delta^{\max}$ should therefore be designed by the final user, depending on the type of solution searched for.

$\delta^{\max}$ **(minutes)**	**30**	**45**
Share of trains deviated (%)	73.7	69.4
Share of flights deviated (%)	15.04	13.3
Train average deviation (minutes)	15.2	16.5
Flight average deviation (minutes)	12.3	13.3

Table 3.8 – Average leg deviation, and share of legs deviated, for the resolution of the ARSTS on the passenger demand instance 1 of December 2, 2019 for scenario S3, for two values of the parameter $\delta^{\max}$. The average deviation is computed among the legs deviated from their initial scheduled departure time.

As a general observation, as the value of $\delta^{\max}$ increases, the increase in total schedule deviation reduces. This is explained by the presence of the station capacity constraints. Indeed, our model implies that the ideal situation for passengers would be that all trains arrive at the same time at the train station, and that all flights depart between $\underline{t_\theta}$ and $\overline{t_\theta}$ minutes later. However, train stations and airports cannot schedule all trains and flights at the same time. Therefore, even if the value of $\delta^{\max}$ gets very large, the transfer quality remains limited by the station capacities. In addition, one can assume that the specific time of the day has an impact on the passenger travel demand : some passengers wish to arrive in the morning while others prefer to reach their destination in the evening, etc. It is therefore preferable for transportation providers to schedule several services at different times of the day in order to satisfy a higher number of passengers.

3.1.4 Strategic timetable synchronisation problem conclusion

A metric has been proposed to measure the level of synchronisation between air and rail timetables. The function allows one to provide a score, depending on the connecting time between trains and flights, that takes into account station processing time, and eventual transfer time between legs. Based on this metric, a MILP formulation of the strategic air-rail timetable synchronisation problem at the network scale was introduced, with the purpose of providing smooth connections for passengers transferring between trains and flights, at a limited cost for transportation service providers. This optimisation considers operational constraints such as airport and train station capacities, and the guarantee of air-air connections for airlines. In order to assess the benefit of air-rail collaboration, the model was solved at the European scale across three countries on three European airport hubs : Paris-Charles de Gaulle, Frankfurt and Madrid-Barajas airports. Three scenarios were proposed, in which each operator (airlines and railway companies) agrees to change its schedule or not. Results show first that a collaboration between rail and air can yield a gain of 5.5 minutes and 9.1 minutes per connection for passengers, in 2019 and 2021, on average. In addition, the number of short connections is decreased by 46% and 56% in 2019 and 2021, respectively, reducing thereby the risk of missed connections for passengers. At the same time, initially long connections are shortened, reducing the wait time of passengers at airports. These results were obtained by changing the schedule of less than 16% of the flights operated over Europe, and 74% of the trains, in 2019. Remember that only a subset of operating trains (the ones stopping at one of the three considered trains stations) were considered here, reducing the actual share of trains deviated.

Several avenues for future work are under consideration. First, the robustness of the new schedule could be assessed against several delay patterns. Indeed, delays may severely impact the smoothness of a door-to-door journey, especially in case of missed connections. On the studied period, the week of December 2019, the average delay observed in the historical data is 10 minutes. This information could be included in the optimisation process to ensure smooth connections, even in case of disruptive events. Then, the set of re-accommodation options for passengers in case of missed connections could be included in the optimisation model. Indeed, missing a connection with a flight operated several times a day is less critical than with a daily flight. A further sensitivity analysis could be lead to assess the impact of the model parameters ($\mathrm{MCT}_\theta, \mathrm{MACT}_\theta, \underline{t_\theta}, \overline{t_\theta}$, Δ, r, $c_{\theta,1}$, $c_{\theta,2}$) on both the resolution performances and the solutions obtained. Moreover, piecewise linear functions with additional breakpoints, or even quadratic cost functions should be tested. Indeed, the disutility of *really* short connections (whose connection time is close to MCT_θ) is in practice higher than that of connections whose the connection time is close to $\underline{t_\theta}$. Therefore, it may be relevant to improve these connections in priority. Finally, schedule elasticity should be considered since changing train or flight departure times is likely to affect the demand.

3.2 Tactical planning horizon

In the previous sections, an air-rail schedule synchronisation methodology was developed to provide passengers with smooth transfers between legs. The main objective was to create an integrated planning between trains and flights that both let a sufficient amount of time for passengers to transfer between legs, even in the case of delays, and reduces additional wait time at stations. Unfortunately, it may happen that delays on the first leg are too large and that passengers still miss their connection. In addition, such integrated timetables are not established today. In this section, we propose to adapt our optimisation model in order to create a synchronisation framework at the tactical planning horizon. When passengers travel, delays may occur on one leg of their journey, threatening connections with the next ones. For example, if a train or a plane is delayed, passengers connecting with a flight may hope that the latter is also delayed in order to maintain their connection. In practice, depending on airline strategies, some airlines may wait for delayed passengers travelling within the same airline. However, such a mechanism is not always proposed, especially for ground transportation modes. In this section, we propose to assess the benefit of such an air-rail collaboration in presence of delays, both for passengers and transportation service providers. The model developed at the strategic level is adapted, and a new passenger-oriented metric, focusing on the tactical level is proposed. More precisely, the objective of the study is to delay targeted trains or flights in order to wait for delayed passengers and avoid missed connections. This work was made in collaboration with Geoffrey Scozzaro, also PhD student at ENAC, and was published in Scozzaro *et al.* (2023).

Previous works have studied the Delay Management (DM) problem on the ground side. Schöbel (2001) is the first author to formulate the problem of deciding whether or not to delay a vehicle in a public transport system to wait for transferring passengers. She proposes a mixed-integer formulation to minimise the total delay of passengers at their final destination. She assumes that the passenger's delay is equal to the delay of his train, if he catches it, or to some predefined constant otherwise. Later, Schöbel (2009) considers track capacity constraints for a railway system. Dollevoet *et al.* (2012) propose integrating passenger rerouting into the DM process. The same authors consider station capacity constraints and track re-allocation in Dollevoet *et al.* (2015). For a review of DM problem handling, the reader can refer to König (2020).

On the airside, Santos *et al.* (2017) are the first to propose a version of the delay management problem applied to the airline. Montlaur and Delgado (2017) consider the problem of balancing airport capacity at a hub airport by assigning delays to departing flights at a pre-tactical level (a few days before operations) and to arriving flights at a tactical level (the day of operations). They test different strategies to minimise either flight or passenger delays, considering connecting passengers and turnaround constraints. Delgado *et al.* (2016) propose a delay recovery strategy at a hub airport through gate delays to wait for delayed connecting passengers and a dynamic cost index to recover from such delays. Delgado *et al.* (2021) propose an agent-based model for handling air traffic delays

through 4D trajectory adjustment to reduce costs and delays for connecting passengers.

Collaboration between air and ground transportation systems received a growing interest over the past few years. Li *et al.* (2018) present an overview of actual collaboration between airlines and train service providers to create an integrated air-rail service for passengers. Laplace *et al.* (2014) present the META-CDM project, which aims at involving ground transport stakeholders into the airport Collaborative Decision Making (CDM) system to improve passenger door-to-door journeys. In this context, studies on multimodal recovery solutions in case of massive disruptions show promising results in mitigating the impact of such events on passengers (see for instance Dray *et al.* (2015); Marzuoli *et al.* (2016, 2015)). Scozzaro *et al.* (2022) propose flight rescheduling at the tactical level to mitigate the impact of airport access mode disruptions on passengers. They consider airside constraints such as terminal capacity, maximum runway throughput, or minimum passenger connecting time. Their work focuses on a single airport and does not consider reactionary delay.

Here, we propose a tactical delay management strategy at the network level, combining the works of Buire *et al.* (2022) and Scozzaro *et al.* (2022). This study is the first to address the delay management problem in a long-haul multimodal network, combining constraints on the air and ground sides. We extend the original version of the problem developed by Schöbel (2001). We take into account real operational constraints, such as airport and railway station capacities, Air Traffic Flow Management (ATFM) slot adherence, or even minimum aircraft turnaround time. We also consider the reallocation time for passengers who miss their connections.

3.2.1 Total passenger delay metric

As explained by Cook *et al.* (2012), flight delays do not necessarily capture the actual delays experienced by passengers. The situation is similar for train delays, which can lead to missed connections, and potentially late arrivals at the final destination. We therefore introduce the *total passenger delay* metric, as the sum of the delays experienced by passengers when arriving at their final destination. To compute passenger delays, we propose to partition passengers into three different groups :

— *on-time passengers* : passengers who catch their flight/train ; their delay is equal to the delay of the flight/train ;

— *reallocated passengers* : passengers who miss their connections due to a delay on the first leg, they are consequently reallocated to another flight/train going to the same destination within the same day ;

— *stranded passengers* : passengers who miss their connections and remain without reallocation option (no seat available or no more flight/train going to the same destination within the day).

The delay of *reallocated passengers* is computed as follows. For each flight and train, we consider the direct alternative, enabling the passengers to arrive at their destination with the smallest possible delay. This alternative can be either a train or a flight. In this study, we only consider direct alternatives for the sake of simplicity. The delay of *reallocated passengers* corresponds to the difference between the arrival time of the new flight/train at the destination and that of the initial one. Regarding *stranded passengers*, remark that in practice, the reallocation process is a complex task, depending on airline seat availability and strategy. Passengers may be reallocated to the first flight to the same destination in the next morning or be re-routed via another airport to reach their final destination within the day. For simplicity, these options are not considered here and we simply assume that stranded passengers will be re-accommodated to the same flight on the next day at the same departure time, thereby experiencing a 24-hour delay.

The objective of the problem is therefore to reschedule flights and trains at the tactical level to minimise the total passenger delay.

3.2.2 Problem description and mathematical formulation

In the event of disruptions on the ground or on the air sides leading to train or flight delays, we assume that service providers are notified ahead of time about the affected vehicles and their expected delays for the remainder of the day. For instance, consider a power outage on a railway network between 6am and 8am, causing delays for several trains throughout the day due to a domino effect. We assume that the rescheduling of trains and flights can occur once operators anticipate delays caused by the incident, such as when power is restored at 8am. The key challenge is deciding whether a vehicle should wait for connecting delayed passengers. For example, consider a flight of 100 passengers scheduled to leave at 9am, with 10 passengers connecting from a previous train. Due to the disruption, these passengers arrive at the boarding gate 10 minutes after the scheduled boarding time. There are two options : depart on time or delay the flight. On the one hand, if there is another flight to the same destination in three hours, departing on time will result in a total passenger delay of $3 \times 60 \times 10 = 1,800$ minutes. On the other hand, if the flight waits for the delayed passengers, the total passenger delay will only be $100 \times 10 = 1,000$ minutes. In this situation, the aircraft should wait for the connecting passengers. However, if only five passengers were connecting, it would be better to depart on time.

Assuming that delays arise on several trains or flights during the day, the problem consists in assigning tactical delays to other trains and flights so as to minimise the total passenger delay.

3.2.2.1 Modelling network and passenger itineraries

In the following, the network model and the notations introduced in Section 3.1 are used. In addition, several operational constraints previously defined are also considered

here :

- the number of trains scheduled to stop at each train station cannot exceed the number of tracks at this station (train-station capacity constraint) ;
- the number of airport departure and arrival movements, operated every 10-minute and 60-minute intervals is limited (airport capacity constraint),
- a minimum turnaround time between two flights operated by the same aircraft is considered ;
- the train dwell time at the station must remain the same as in the initial schedule ;
- the train and flight travel times remain unchanged ; (each vehicle maintains its scheduled speed).

In this study, we also assume that the passenger demand (previously denoted $\mathcal{E}$) is known.

Here, we will better refer to this demand as passenger itineraries, since it is a real-time operation and passengers are already travelling. In addition, these itineraries will be modelled differently from the previous section.

The total delay is computed among all passengers, either *direct* or *connecting* passengers. A direct passenger makes a point-to-point trip, without connecting between two legs at an intermediate station. A connecting passenger uses at least two different legs to travel, and connects at an intermediate station. We therefore introduce, for each leg $l \in \mathcal{L}$, v_l^{direct} to denote the passenger volume of direct travellers of leg l, and v_l^{connect} to denote the passenger volume that connects to/from leg l. In addition, for two legs $l, l' \in \mathcal{L} \times \mathcal{L}$, let $v_{l,l'}$ denote the passenger volume transferring from leg l to leg l'. Figure 3.18 illustrates these notations.

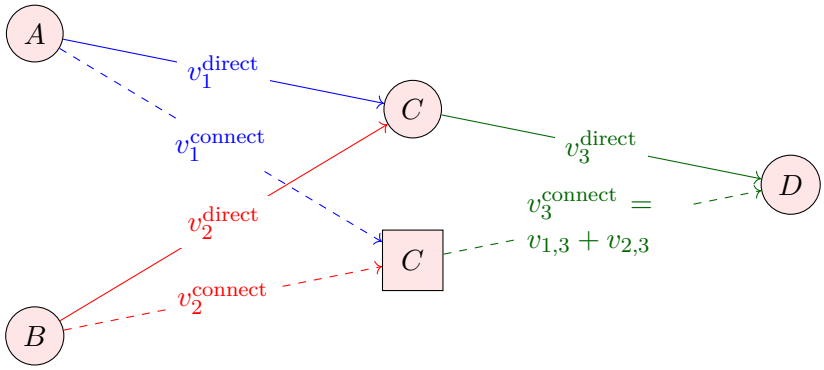


Figure 3.18 – Illustration of direct (plain arc) and connecting (dashed arc) passengers of three legs (blue, red and green) between stations A, B, C, and D. Circle nodes correspond to artificial departure or arrival stations, and rectangle nodes to artificial transfer areas at the station.

For each leg $l \in \mathcal{L}$, let C_l denote the set of legs whose at least one passenger connects with leg l. For example, suppose two passengers arrive at the airport with a flight l_1, and

another one with a flight l_2, and that these three passengers connect with the flight l_3, then $C_{l_3} = \{l_1, l_2\}$.

3.2.2.2 Real-time operational constraints

Several additional constraints related to our real-time context are considered. First, the model we are about to propose should mitigate the impact of a disruption on passengers affected by such an event, but such a mitigation should not be made at the expense of other passengers. Therefore, our model should ensure that passenger having a connection whose first leg was initially on time still have their connection. More precisely, the model should ensure that these passengers benefit from a MCT to connect. Second, in order to manage the airspace capacity, the departure time of aircraft may be regulated through so-called Air Traffic Flow Management (ATFM) departure slots. These slots are allocated by a central unit, the European Network Manager, upon the request of the local Flow Management Position (FMP), when an imbalance between demand and capacity is foreseen at airports and/or en-route. Aircraft subject to the ATFM slot management should therefore depart between [-5,10] minutes around their scheduled departure time. The model must therefore constrain the departure time of regulated aircraft to be lower than Δ_{ATFM} minutes after the initial scheduled departure time. In the sequel, we call this constraint is referred to as the *ATFM slot adherence* constraint. Finally, in order to mitigate the impact on the transportation network, we limit the delay that could be assigned to flights and trains to wait for disrupted passengers to Δ minutes.

3.2.2.3 Optimisation problem formulation

This subsection details the input data, the decision variables, the objective function, and the constraints of the optimisation problem.

3.2.2.3.1 Input data Similarly to Section 3.1.2.4.1, the airport and train station capacities are known, as well as the initial scheduled arrival time, T_l^A, and departure time, T_l^D, of each leg $l \in \mathcal{L}$. For each leg $l \in \mathcal{L}$, the number of direct passengers v_l^{direct}, and the set of legs with connecting passengers C_l, are known. In addition, for each leg $l \in \mathcal{L}$, for each leg $l' \in C_l$ with connecting passenger, the volume of connecting passenger $v_{l',l}$ is given as input data. For these passengers, the reallocation delay estimator is known and denoted r_l. As detailed above, some passenger connections must be maintained as they are not directly impacted by the disruption. Let C^p denote these priority connections, which are assumed to be known. Finally, let $\mathcal{L}^{\text{ATFM}}$ denote the set of aircraft subject to the ATFM slot adherence.

3.2.2.3.2 Decision variables For each leg $l \in \mathcal{L}$, we define the integer decision variable d_l, the delay assigned to leg l. We further define the auxiliary integer decision variables k_l^D

and k_l^A, the new index departure time step and arrival time step of leg l, and the discrete variables t_l^A and t_l^D, the new scheduled arrival time and departure time of l, $l \in \mathcal{L}$. For each passenger connection, $l \in L, l' \in C_l$, let $y_{l',l}$ be the binary decision variable indicating whether the connection is still feasible or not. The decision variable $\delta_{l',l}$ measures the delay experienced by passengers connecting from leg l' to leg l, for $l \in \mathcal{L}, l' \in C_l$. Finally, similarly to Section 3.1.2.4.2, for each time step $i \in \mathcal{T}$, and each leg $l \in \mathcal{L}$, let $x_{l,i}^D$ and $x_{l,i}^A$ be binary decision variables indicating whether leg l is scheduled to depart after time step i, and scheduled to arrive after time step i, respectively.

Notations newly introduced for the tactical synchronisation problem are summarised in Table 3.9.

Table 3.9 − Notations of the air-rail tactical timetable synchronisation problem

Sets	
$\mathcal{L}^{\text{ATFM}}$	index set of flights subject to the ATFM slot adherence constraint
$\mathcal{C}^{\text{p}}$	index set of *priority* leg pairs (for which passenger connections must be maintained).
$\mathcal{C}_l$	index set of legs with passengers connecting to leg l, $l \in L$
Input data	
Δ	maximum pushback parameter, multiple of h
Δ_{ATFM}	maximum pushback parameter for flights subject to the AFTM slot adherence constraint, multiple of h
$\text{MCT}_{l',l}$	minimum connection time to connect from leg l' to leg l, $l \in L, l' \in \mathcal{C}_l$
v_l^{direct}	volume of passengers using l as a direct connection, $l \in L$
$v_{l',l}$	volume of passengers transferring from leg l' to leg l, $l \in L, l' \in C_l$
r_l	reallocation delay for passengers missing their connection with leg $l, l \in L$
Decision variables	
d_l	tactical delay assigned to leg l, $l \in L$ (in minutes)
$\delta_{l',l}$	delay experienced by passenger connecting from leg l' to leg l, $l \in L, l' \in \mathcal{C}_l$ (in minutes)
$y_{l',l}$	binary, indicates whether the connection from leg l' to leg l is still feasible or not, $l \in L$, $l' \in \mathcal{C}_l$

3.2.2.3.3 Objective function and constraints The proposed model reads as follows :

$$\min_{k,t,x,d,\delta,y} \sum_{l \in L} \left(v_l^{\text{direct}} d_l + \sum_{l' \in C_l} v_{l',l} \delta_{l',l} \right) \tag{3.4}$$

subject to :

$$t_l^D = h(k_l^D - 1) \qquad l \in \mathcal{L} \qquad (3.4a)$$
$$t_l^A = t_l^D + IVT_l \qquad l \in \mathcal{L} \qquad (3.4b)$$
$$k_l^A = k_l^D + \frac{IVT_l}{h} \qquad l \in \mathcal{L} \qquad (3.4c)$$
$$k_l^D \leq i + Mx_{l,i}^D \qquad l \in L, i \in \mathcal{T} \qquad (3.4d)$$
$$\sum_{i=0}^{|\mathcal{T}|-1} x_{l,i}^D = k_l^D \qquad l \in \mathcal{L} \qquad (3.4e)$$
$$k_l^A \leq i + Mx_{l,i}^A \qquad l \in L, i \in \mathcal{T} \qquad (3.4f)$$
$$\sum_{i=0}^{|\mathcal{T}|-1} x_{l,i}^A = k_l^A \qquad l \in \mathcal{L} \qquad (3.4g)$$
$$t_l^D \leq t_l^{D,0} + \Delta \qquad l \in \mathcal{L} \qquad (3.4h)$$
$$t_l^D \leq t_l^{D,0} + \Delta_{\text{ATFM}} \qquad l \in \mathcal{L}^{\text{ATFM}} \qquad (3.4i)$$
$$t_l^D - t_l^{D,0} = d_l \qquad l \in \mathcal{L} \qquad (3.4j)$$
$$t_{l_2}^D - t_{l_1}^A \geq MTAT_{l_1,l_2} \qquad (l_1, l_2) \in \mathcal{P}^{\text{air}} \qquad (3.4k)$$
$$t_{l_2}^D - t_{l_1}^A = DW_{l_1,l_2} \qquad (l_1, l_2) \in \mathcal{P}^{\text{rail}} \qquad (3.4l)$$
$$\sum_{l \in \mathcal{L}_n^D \cap \mathcal{L}_n^A} (x_{l,i}^D - x_{l,i}^A) \leq O_n^{\max} \qquad n \in \mathcal{N}^{\text{rail}}, i \in \mathcal{T} \backslash \{0\}, \qquad (3.4m)$$
$$\sum_{\tau=i}^{i+w-1} \sum_{l \in \mathcal{L}_n^A} x_{l,\tau}^A - x_{l,\tau+1}^A \leq Y_n^{A,w,i} \qquad n \in \mathcal{N}^{\text{air}}, w \in W, i \in \mathcal{T}_w \qquad (3.4n)$$
$$\sum_{\tau=i}^{i+w-1} \sum_{l \in \mathcal{L}_n^D} x_{l,\tau}^D - x_{l,\tau+1}^D \leq Y_n^{D,w,i} \qquad n \in \mathcal{N}^{\text{air}}, w \in W, i \in \mathcal{T}_w \qquad (3.4o)$$
$$t_{l'}^D - t_l^A \geq \text{MCT}_{l',l} \qquad (l, l') \in \mathcal{C}_l^{\text{p}} \qquad (3.4p)$$
$$t_{l'}^D - t_l^A + My_{l',l} \geq \text{MCT}_{l',l} \qquad l \in L, l' \in \mathcal{C}_l \qquad (3.4q)$$
$$\delta_{l',l} \geq y_{l',l} r_l \qquad l \in L, l' \in \mathcal{C}_l \qquad (3.4r)$$
$$\delta_{l',l} \geq d_l \qquad l \in L, l' \in \mathcal{C}_l \qquad (3.4s)$$
$$y_{l',l} \in \{0, 1\} \qquad l \in L, l \in \mathcal{C}_l \qquad (3.4t)$$
$$x_{l,i}^A, x_{l,i}^D \in \{0, 1\} \qquad l \in L, i \in \mathcal{T} \qquad (3.4u)$$

where d is a vector whose l^{th} component is d_l, and y and δ are matrices whose (l', l) components are $y_{l',l}$ and $\delta_{l',l}$, respectively. The model aims at minimising passenger delay across the air and rail networks. Constraints (3.4a) and (3.4b) link the new departure time step to

the departure-time and arrival-time variables of leg l. Constraints (3.4c) link the arrival time step of leg l with its departure time step. Constraints (3.4d) and (3.4e) link the values of the binary decision variables $x^D_{l,i}$ witht the values of variables k^D_l. Similarly, constraints (3.4f) and (3.4g) link the values of the binary decision variables $x^A_{l,i}$ with the values of variables k^A_l. Constraints (3.4h) limit the delay assigned to each leg l. Constraints (3.4i) ensure ATFM slot adherence. The delay assigned to each leg is linked to variables t^D_l via constraints (3.4j). The minimum turnaround time constraints and the constant train dwell time constraints are given by equations (3.4k) and (3.4l), respectively. Constraints (3.4m) stipulate that the number of trains scheduled to stop at a station does not exceed the number of tracks. Constraints (3.4n) and (3.4o) enforce upper bounds on arrival and departure flight movements per time window, respectively. Constraints (3.4p) ensure that passenger minimum connecting times for priority connections are maintained. Constraints (3.4q) link the value of variables $y_{l',l}$ that characterise if passengers connecting between legs l' and l miss their connection or not, with variables t^D_l and t^A_l. Constraints (3.4r) and (3.4s) fix the reallocation delay between flights l' and l to d_l if passengers have their connection, and to the reallocation delay r_l, otherwise. Finally, constraints (3.4t) and (3.4u) define the definition domain of the decision variables. Remark that, similarly to Problem (3.2), for each leg $l \in \mathcal{L}$, decision variables k^D_l and k^D_l are automatically integer variables since they are the sum of binary decision variables.

3.2.3 Computational experiments on the Western Europe case study

This section focuses on the Western Europe case study. It first outlines the data used and the assumptions made. It then describes in detail the post-processing procedure for reallocating passengers, which is crucial for accurately assessing the total passenger delay. Finally, numerical results are presented and discussed.

3.2.3.1 Western Europe case study

This case study focuses on the historical day of December 4, 2019 when the French National Railway Providers (SNCF) went on strike. We gather initial flight schedules (Eurocontrol, 2023) from the 18 largest airports in France, Germany, and Spain, including three major hub airports : Frankfurt airport (FRA), Madrid-Barajas airport (MAD), and Paris-Charles de Gaulle airport (CDG). Throughout this day, 10,407 flights were operated, with 593 departures scheduled at CDG. It is important to note that we could not access either the actual train delay data nor the number of train cancellations. According to SNCF (2023b), accross December 2019, the high-speed-rail punctuality in France was around 85%. Therefore, we simulate the disruption by arbitrarily delaying 30% of trains arriving at CDG high speed rail train station, twice the average delay for the month. The delay times are randomly selected using a uniform distribution ranging from a minimum delay, denoted as $t_{\min}$, to a maximum delay, denoted as $t_{\max}$.

Table 3.10 – Case study characteristics.

Case study description	
Case Study	Western Europe Transport Network
Number of airports	496
Number of train stations	72
Number of flights	10407 (593 from CDG)
Number of trains	646 (66 to CDG station)
Airports with limited capacity	18 largest airports in France, Germany, and Spain
Airport with connecting passengers	CDG, FRA, MAD
Train stations with limited capacity	3 stations, each associated with a hub
Train schedule data source	GTFS data
Flight schedule data source	OAG
Minimum aircraft turnaround time (TAT)	45 min
Disruption scenario characteristics	
Date	4 December 2019
Considered events	French railway company on strike
Disruption duration	From 00 :00 to 23 :59
Train delay percentage	30% of trains are late at CDG
Train delay duration (min)	$X \sim \mathcal{U}(30, 90)$, (*i.e.*, $t_{\min} = 30$ min and $t_{\max} = 90$ min)
Train cancellation	Not considered
Flight/Train travel time	Constant
Priority flights	25% of flights need to comply with their ATFM slots at main airport hubs
ATFM delays	Not considered
Maximum priority-flight delay (Δ_{ATFM})	10 min
Maximum flight delay (Δ)	30 min

ATFM delays are not considered here. Therefore, *delayed* flights are only those impacted by the rescheduling methodology that we introduced in this chapter. The maximum delay assignable to a flight, Δ, is set to 30 minutes, and we assume that 25% of flights at each main hub airport were subject to ATFM slot adherence. This percentage is arbitrarily fixed and can be tuned by a final user, depending on the specific airport characteristics. The maximum assignable delay for flights subject to ATFM slot adherence, Δ_{ATFM}, is set to 10 minutes.

Additionally, we allow train tactical rescheduling to ensure compliance with the maximum train station capacity, which may have been compromised due to the initial train

Table 3.11 – Number of connecting passengers per airport. A distinction is made between train-air connections and air-air connections.

Connection type	CDG	FRA	MAD
Air-Air	30,638	41,599	22,277
Train-Air	18,022	12,101	8,168

delays and disruptions caused by the strike. Lastly, we assume that all information regarding train delays and connecting passengers is fully known before running our rescheduling method. Therefore, we employ a one-iteration process to reschedule all legs operated from the morning until the end of the day. Table 3.10 summarises the characteristics of the case study considered.

The number of connecting passengers is simulated after following the methodology developed in Section 2.2 of Chapter 2, and the values are presented in Table 3.11.

Similarly to the rebooking procedure introduced by Ball *et al.* (2010), we propose the following passenger reallocation procedure. Recall that our mathematical model assumes that passengers will be accommodated on the next flight to the same destination if they miss their scheduled flights, although each aircraft has a finite capacity, defined by the number of seats it can offer. To overcome this limitation, we present a post-processing method that effectively reallocates stranded passengers to other flights, taking into account aircraft capacity. Since we do not know the actual number of seats available, we simply assume an 80% load factor for each aircraft. For example, if the historical data recorder that a flight carried 50 passengers, we assume that it had $50 \times \frac{100}{80} - 50 \approx 12$ available seats. We extend this reallocation approach also to direct trains as an alternative re-accomodation option for passengers, again assuming an 80% load factor for each train, as no information is available.

The reallocation process follows a systematic sequence. We consider the chronological list of passengers who have missed their flights and a corresponding set of feasible direct alternatives for each individual. These alternatives are ranked according to the delay they cause at the passenger final destination. For each passenger, we offer the best available re-routing option (in terms of delay). In the case where a passenger's best alternative flight/train is full, we select its second best option, and so on. When no re-accommodation option is available, the passenger is stranded, and he is subject to the (arbitrary) 24-hour delay. Note that this post-processing reallocation procedure is operated after the rescheduling (optimisation methodology), *i.e.*, assigned flight and train delays are considered in the re-allocation procedure.

3.2.3.2 Numerical results

Again, computations are performed using an AMD Ryzen 5 4500U CPU and 16 GB RAM laptop. The resolution of the optimisation problem formulation is made with the MIP solver Gurobi, version 9.1.2 (Gurobi Optimization, LLC, 2023). The computation time is 23 seconds.

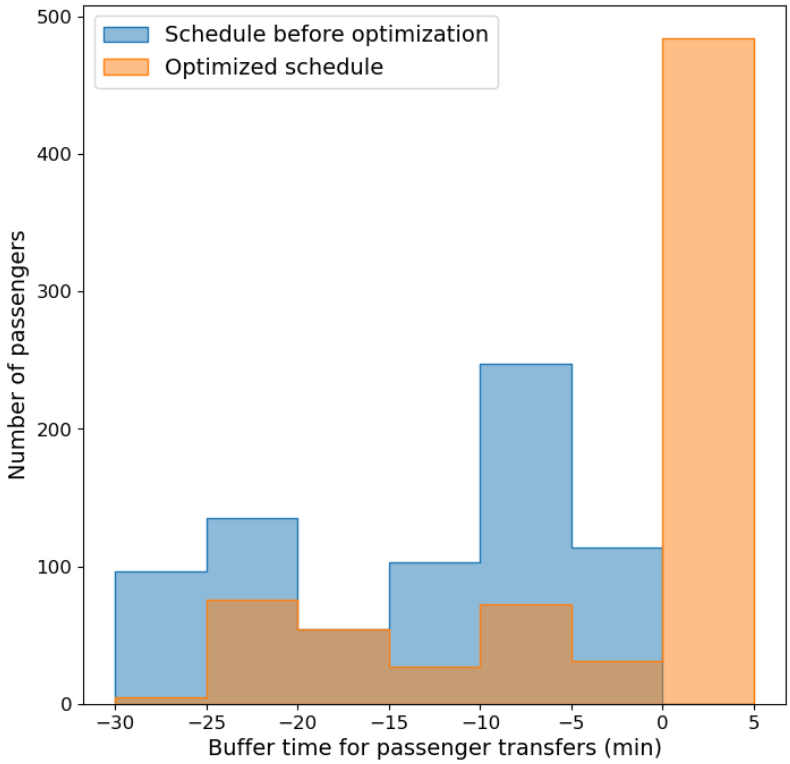


Figure 3.19 – Distribution of passenger transfer buffer times before and after rescheduling. Buffer times are calculated by subtracting the minimum connecting time from the actual passenger transfer time. A negative buffer time indicates that passengers do not have enough time to transfer, caused by a delay on their first leg. This graph only shows passengers who could recover their initial flights thanks to the rescheduling (*i.e.*, missing their flights by 30 minutes or less before rescheduling).

We call *transfer buffer time* the difference between the actual passenger transfer time and the minimum required connection time. Figure 3.19 displays the distribution of buffer time for passengers transferring from a train to a flight at CDG airport. Only buffer times of passengers who would have missed their flight based on the original schedule but can still make it on time if the flight is delayed, are displayed. Passengers who arrive before the initial

departure time or who arrive more than 30 minutes after the initial departure time are not presented in the figure. The figure shows a significant increase in passenger connections with a 0-minute buffer time after rescheduling. A 0-minute buffer time corresponds to a transfer time equal to the minimum connection time required for passengers to catch their flight. Consequently, the delay management strategy allows 484 of the 1,221 passengers who initially missed their flights to arrive on time for boarding. The rescheduling does not induce buffer time strictly larger than 0 minutes for these passengers as this would delay the *on-time passengers* and, therefore, increase the total passenger delay.

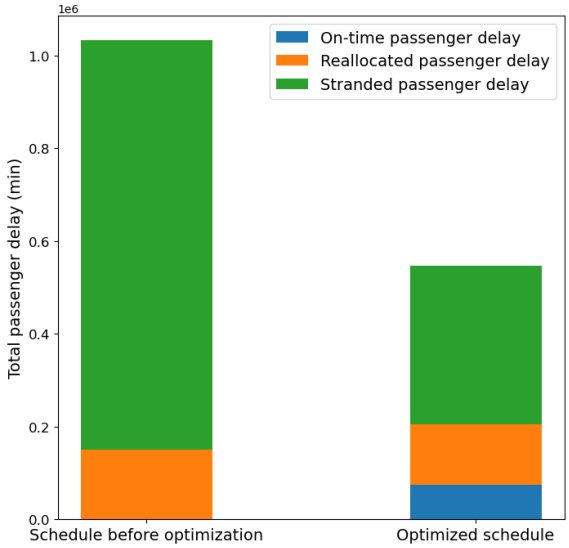


Figure 3.20 – Total passenger delays before and after rescheduling, stacked by passenger types (on-time passengers, reallocated passengers and stranded passengers).

Figure 3.20 depicts the total delay experienced by passengers before and after optimisation. The main difference between the initial and the optimised schedule lies in the number of stranded passengers. Indeed, 614 passengers have no reallocation option before optimisation and would have to wait until the next day to reach their final destination. After optimisation, the number of stranded passengers is reduced by 71%, and the total passenger delay by 55%. Indeed, the algorithm prioritises these passengers if the flight can wait since the cost of a missed connection is high. However, the maximum flight delay authorised to wait for passengers is 30 minutes (or 10 minutes for priority flights that need to respect their departure slots). Hence, some passengers might not have their connections if the required time to make the connection is above that limit. Therefore, several passengers remain stranded even after the rescheduling. Finally, the total delay experienced by direct passengers departing from CDG is 12,810 minutes, resulting in an average passenger delay of 0.3 minutes. As a result, the rescheduling has a marginal impact on passengers whose

train or flight arrived on time.

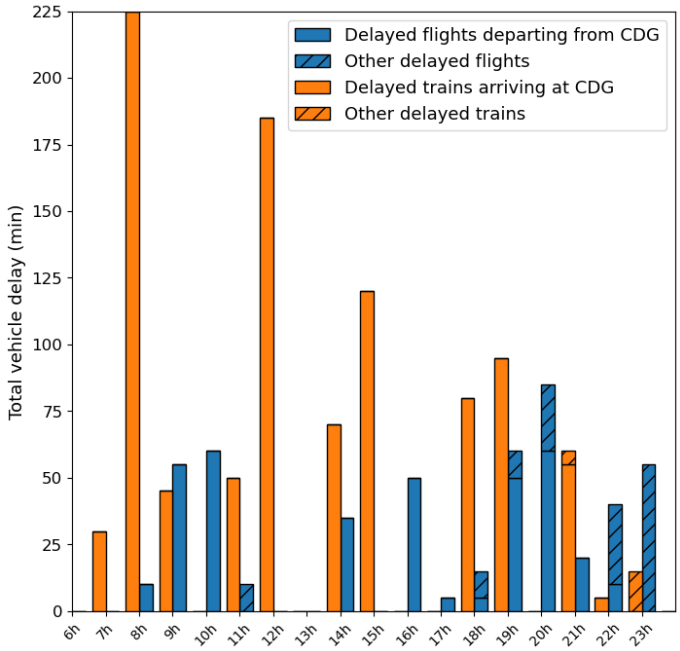


Figure 3.21 – Distribution of total vehicle delays, after optimisation, per hour. Flight and train delays are displayed in blue and orange, respectively. The hatched bars represent propagated delays.

Regarding operator delays, Figure 3.21 displays the total vehicle (train or aircraft) delay per hour. Orange plain bars represent the total train delay, including the delay due to the strike and the one assigned during rescheduling due to train station constraints. One observes that most of the delayed trains are in the morning. The hatched bars correspond to flights not departing from CDG and trains not arriving at CDG, *i.e.*, the reactionary delay on the network. Reactionary delays may occur for several reasons. A tight initial turnaround time cannot absorb the delay of an arriving flight. Limited airport and train station capacities can also lead to rescheduling other flights and trains to avoid congestion. Finally, delayed flights or trains with connecting passengers can also create reactionary delays at their arrival station to maintain passenger connections. After rescheduling, seven trains are delayed, including four at stations other than CDG. 35 flights are also delayed, among which eight are not from CDG. Significant flight delays are observed during the morning rush hour (9am and 10am) and the evening (7pm and 8pm). The morning hours see a surge in missed passenger connections due to significant train delays in the previous hour. The second peak of flight delays is either due to reactionary delays from previous flights (displayed by hatch bars) or fewer flight reallocation options. Indeed, passengers

who miss their connections at the end of the day are more likely to be stranded without reallocation options until the next day. As the cost of these missed connections is high, the rescheduling algorithm delays these flights in priority so as to wait for passengers.

On average, due to the rescheduling, all flights across Europe experience a delay of 2.4 seconds, while the departing flights at CDG experience a delay of 50 seconds. The proposed rescheduling plan delays 5% of the departing flights at Paris-CDG airport by 13 minutes on average. To put this into perspective, Table 3.12 shows the characteristics of the actual delays experienced by flights during the historical operating day in question. As

Table 3.12 – Actual flight delay on December 4th 2019 (in minutes) (source : Eurocontrol).

	CDG	ALL
Average actual flight delay	11.0	6.1
Maximum actual flight delay	120	140

per the table, departing flights at CDG were operated with an average delay of 11 minutes. Therefore, our proposed rescheduling approach seems reasonable when compared to the actual delays the airport must face with during a typical operating day.

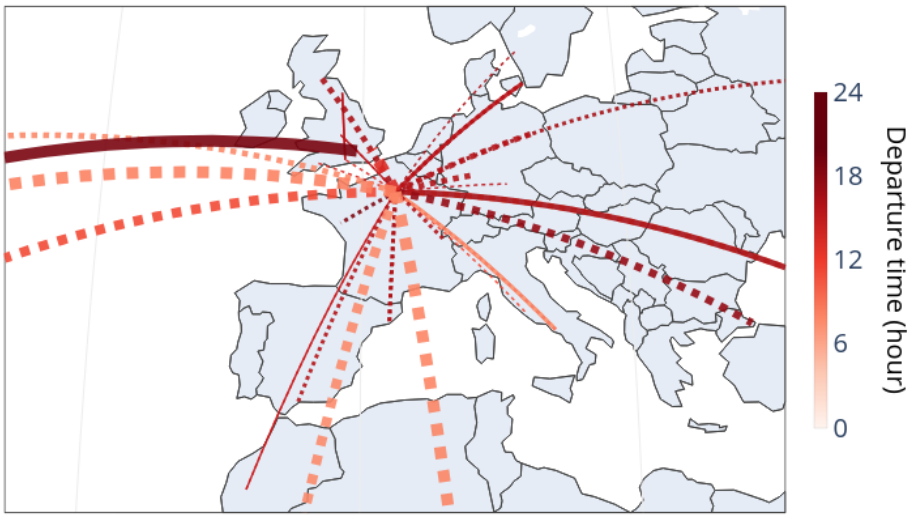


Figure 3.22 – Visualisation of post-rescheduling flight delays. The linewidth and the colour-coding system indicate the delay magnitude and the departure time of the day, respectively. Dotted-line arcs correspond to flights departing from CDG airport, plain-line arvs to other flights.

Figure 3.22 shows a map of delayed flights and the magnitude of these delays. The colour and the width of the arcs correspond to the departure time and the delay assigned to the

flight, respectively. More specifically, a darker colour indicates that the flight's departure time is later in the day, and the wider the arc is, the greater the delay is. Note that long-haul flights are generally assigned the highest delays. The colour of the arcs representing these flights also indicates that they are scheduled in the morning. In fact, these long-haul flights tend to be planned at a daily frequency contrary to short-haul flights. As a consequence, the re-routing time for passengers who miss their connections is 24 hours. On the other hand, delays on short-haul flights are generally assigned in the evening, when passengers have no more opportunities for re-routing. One also observes that a few flights are delayed due to network propagation. These delays occur because the turnaround time initially planned by the airlines between a delayed flight at CDG and the following flight is small. Remark that the rescheduling delays one evening flight departing from the US, by 30 minutes due to the minimum turnaround time constraint and to the assumption of constant in-vehicle time : we assume that airlines do not speed up the flight to recover from delays. However, in practice, the operator would then rather have speeded up the previous flight operated by the aircraft to recover from its departing delay, reducing thereby the impact of the proposed delay management strategy. Taking into account the possibility of such actions is especially relevant for long-distance flights, and could be included in future work. Finally, as mentioned above, exogenous ATFM delays were not considered in this study. However, the proposed rescheduling strategy could deal with these delays by rescheduling flights to wait for connecting passengers while limiting station congestion. Taking these exogenous delays into account will have an impact on the rescheduling solution, as ATFM delays of departing flights reduce the number of passengers missing their rail/air connections.

3.2.4 Conclusion on the tactical timetable synchronisation problem conclusion

Europe's investment in different multimodal research projects underlines the need for collaboration between air and ground transport stakeholders, to provide passengers with reliable journeys. Such air-rail integration would not only improve passenger experience but also allow airlines and airports to have accurate information about passenger connections. This could create a win-win situation for all stakeholders, by boosting passenger demand while limiting extra expenses for the service providers. In this context, we presented a delay management strategy tailored to a large integrated air-rail network. We simulated a disruption occurring on the French railway network that led to passengers missing their connections at CDG airport. The experimental results highlight the effectiveness of our mitigation strategy, demonstrating its ability to reduce passenger delays by 55%, while only delaying 5% of departure flights at CDG airport. By considering the entire network, our delay management strategy creates new flight and train schedules that satisfy operational constraints such as station capacities and minimum aircraft turnaround times at other airports throughout the day. This rescheduling methodology limits the delays spread by identifying which flights are likely to propagate delays.

The research conducted in this section contributes to enhancing the experience of passenger travelling across a multimodal long-distance network. Further research on the operator rescheduling cost and passenger preferences should be conducted to implement the proposed delay management strategy. This extension would ultimately lead to better acceptance among transportation stakeholders, and an improved passenger travel experience. Analysing a potential airside disruption would provide valuable insights into how the computed rescheduling differs. Another interesting extension would be to consider dynamic cost indexing, as proposed by Delgado *et al.* (2016), which relaxes the constant travel-time assumption and allows aircraft and trains to speed up to recover from delays. Finally, the rescheduling process should consider ATFM constraints such as en-route capacity and delays assigned by air traffic controllers to smooth the aircraft flow.

3.3 Conclusion

As previously demonstrated in Chapter 1, transportation stakeholders are reluctant to change their operations, especially to synchronise with other transportation modes. The objective of this chapter was therefore to propose to transportation service providers, synchronisation tools that consider individual operational constraints, and do not significantly impact their operations. Two new optimisation models are proposed to facilitate passenger connections between air and rail. The first method, which relies on a connectivity metric, proposes to alter slightly train and flight schedules, in order to create suitable connections between modes for passengers, at a strategic level. The second method aims at mitigating the impact of a disruption on passengers making intermodal connections. The synchronisation tool proposes to assign tactical delay to trains and flights, so as to wait for passengers impacted by a disruption on their first leg. Both tools have been tested at the level of the Western Europe transport network on realistic instances handling more than 10,000 flights and 400 trains per day, leading to consider more than 100,000 passenger connections. Computational results highlight the potential benefit for passengers, at a limited cost for transportation service providers. Additional features could be included in these two optimisation models and tools. For instance, the daily leg frequency and the time at which passenger connections occur should be included in the strategic synchronisation tool. Indeed, one can assume that it is more important to synchronise connections when passengers have only one choice (no alternative) within the day. Similarly, evening connections should be prioritised as they could lead to missed connections and stranded passengers in case of delays on the first leg. Regarding the tactical synchronisation, ATFM delays should also be included.

These models offer flexibility, enabling individual operators to utilise it independently, possibly taking the other mode into account as a constraint. Alternatively, it can be employed by an external coordinator who centralises data from both operators, suggesting new schedules, as demonstrated in scenarios S1, S2, or S3 for the strategic synchronisation tool. Benefits for passengers can already be demonstrated if only one mode of transport adapts its timetable to the other. Such mechanisms pave the way toward full collaboration between stakeholders as they could already improve the passenger experience, at the expense of a small commitment from transportation service providers. This concept aligns with the newly initiated SESAR project SIGN-AIR (SESAR Joint Undertaking, 2023b). The primary objective of this research initiative is to introduce a coordination platform connecting multiple transportation providers. Each provider decides what data it wishes to share and what contract to be established with other transport providers.

Chapitre 4

Redesigning an air-rail timetable depending on a dynamic demand

In the previous chapter, we propose to improve air-rail passenger transfers by considering marginal adjustments to existing timetables. This chapter goes toward a stronger collaboration, and addresses simultaneously the classical frequency planning problems of both airlines and railway operators. Both the flight and rail frequencies are computed. In addition, we do not assume a hub-and-spoke network topology and we allow direct services when relevant. Passengers' preferences are extracted from the analysis of mobile phone data, and the environmental cost of the transportation network is also taken into account in the objective function. In the following, we will refer to this problem as the Air-Rail Service Network Design (ARSND) problem. The work presented in this chapter results from a two-month internship at NOMMON, a Spanish company specialised in Big-Data applied to mobility.

4.1 Graph model

The transportation network is represented by an oriented graph, $G = (V, A)$, where the vertex set V is the set of nodes, and the arc set A is the set of connections between nodes. The set V is further partitioned into the set of airports, V^{air}, the set of train stations, V^{rail}, and the set of city centres, V^{city}. Similarly, the set A is partitioned into five subsets : A^{air} represents flight routes between airports, A^{rail} corresponds to rail tracks, A^{transfer} models transfers between stations, and A^{access} and A^{egress}, arcs modelling passenger transfers from city centre nodes to transportation stations and passenger transfers from transportation stations to city centre nodes, respectively. In addition, each airport node is duplicated into one arrival node and one departure node, so as to model transfers between two consecutive flights at a same airport. Similarly, a subset of train station nodes (those corresponding to the biggest cities) are also duplicated into one arrival node and one departure node, so as

to model transfers between two consecutive trains at a same train station. An illustration of the considered network is presented in Figure 4.1.

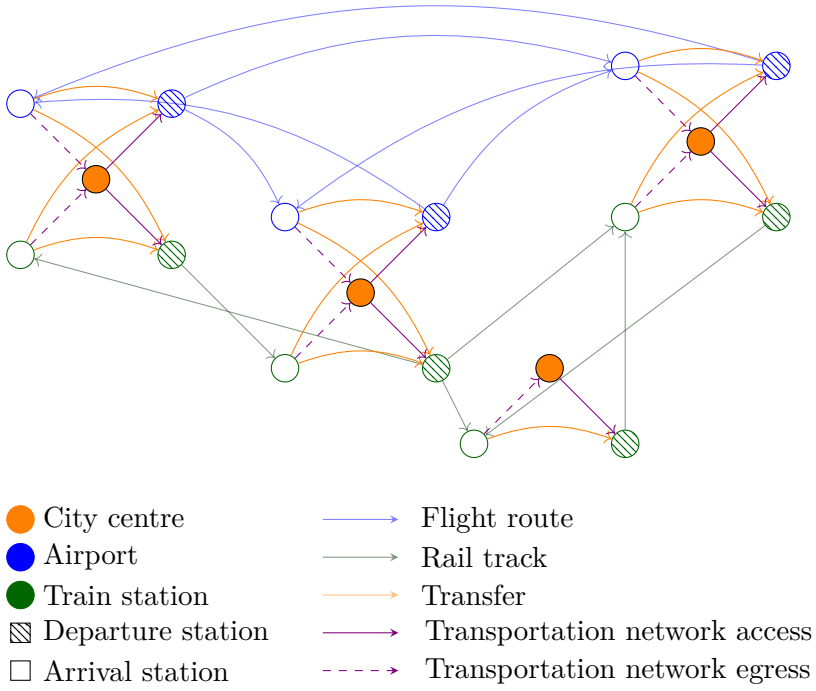


Figure 4.1 – Illustration of air-rail multimodal network : a 4-city example.

For each arc $a \in A$, several costs are defined : the travel time, the price, and the CO_2 emission. They are detailed in the next three subsections.

4.1.1 Travel time cost

For flight-route and rail-track arcs, the travel time is estimated as the average travel time by flight and train initially scheduled by the transportation suppliers. Regarding transfer arcs, an average transfer time, noted t^{transfer}, is considered. If passengers must shift between two stations, the travel time between the two stations is taken into account in the average transfer time. This travel time is computed by dividing the distance between the two stations by an average speed of 90km/h. If the stations are located at the same place (*e.g.*, if the train station is located at the airport), then this transfer time is zero. The travel time from/to the city centre to/from the station is similarly computed for access arcs and egress arcs. Finally, in order to account for the total door-to-door journey duration, station processing times are taken into account within the total travel time. It has been found that airport processing time and passengers conservative behaviour may significantly

increase the door-to-door journey (Innaxis, 2023). We therefore add an average outbound processing time (t^{dep}), to the access arc travel time, and an average inbound processing time (t^{arr}) to the egress arc travel time.

4.1.2 CO_2 cost

The CO_2 cost of travelling by air or by rail is different. Regarding flights, fuel consumption directly depends on the aircraft type, weather conditions, airline strategies, payload (weight of passengers and cargo carried), etc. It is therefore difficult to obtain an accurate measure of the CO_2 emission at such an early stage of the frequency planning process. To overcome this issue, we proceed with the following simplifying assumptions :

— We consider only one possible aircraft type, which can cover all routes in the domestic market.

— We assume that a scheduled flight has all available seats filled or the remaining available payload, else, is filled with freight. Aircraft indeed generally take off at a weight close to its maximal authorised weight according to Sun *et al.* (2018). Thus, we consider an average value of 80% of the maximum take-off weight for each flight.

— According to the aircraft characteristics such as the payload over range mapping, the fuel consumption per kilometre can be estimated as follows : the total amount of fuel (in kg) for a scheduled flight is the product of the consumption per kilometre and the total distance flown.

— Finally, the CO_2 *emission equivalent* is obtained using the International Air Transport Association (IATA) conversion table (International Air Transport Association, 2022).

The CO_2 emission of trains per kilometre is provided by the Office of Rail and Road (Office of Rail and Road, 2022). Therefore, the CO_2 emission of a scheduled train is estimated as the product of the emission per kilometre and the distance travelled by train. For transfers (access and egress arcs), the CO_2 emission is assumed to be null.

4.1.3 Price

The monetary cost of travelling by air or by rail for passengers is estimated as the product of the unit cost of a kilometre and the total distance travelled. Values of unit cost per kilometre for flights and trains are obtained from Livingston *et al.* (2022) and Tauler and Martin (2021), respectively.

4.2 Passenger demand model

The passenger demand corresponds to the number of passengers who want to travel between each Origin-Destination (OD) pair. The OD demand is represented by a set of

commodities represented by an index set C. A commodity $c \in C$ is a tuple (O_c, D_c, d_c), where O_c is the origin node, D_c is the destination node and d_c is the number of passengers for the OD pair (O_c, D_c). For each OD pair, passengers may use several paths to travel. For instance, if two cities are connected by rail and have airports, passengers may use a direct flight or a direct train. For each commodity $c \in C$, let $\mathcal{P}_c$ denote the set of paths that connect O_c to D_c. A *path* p is a sequence of arcs in the transportation graph. Passengers may have various travel preference criteria : the business travellers are likely to be more concerned about the travel duration than leisure travellers are (Pels *et al.*, 2003). Similarly, passengers more concerned about the CO_2 emissions prefer to travel by train. We therefore define a set of cost : $I = \{\text{Travel time, Price, } CO_2\}$. The set of passengers involved in each commodity $c \in C$ has additional attributes, $\alpha_{c,i}$, representing the sensitivity of that group of passengers to the criteria i, $i \in I$. Thus, for each commodity $c \in C$ and for each path $p \in \mathcal{P}_c$, we define τ_{cp} the cost of path p for commodity c as the weighted sum of three costs :

$$\tau_{cp} = \sum_{i \in I} \alpha_{ci} \tau_{pi}, \tag{4.1}$$

where τ_{pi} corresponds to the cost i of path p (*e.g.*, travel time, price and CO_2 emission of p). The cost of a path $p \in \mathcal{P}$ is computed as follows :

$$\tau_{pi} = \sum_{a \in p} c_{ai}, \tag{4.2}$$

where c_{ai} is the cost i of arc $a \in A$, as described above (thus, each arc has three cost).

4.3 Mathematical formulation

The objective of this problem is to determine the daily flight and train frequencies in a domestic market that minimise the passengers' generalised cost and CO_2 emission. This section presents the given input data, the decision variables, the objective function, and the constraints. The resolution approach is detailed in the last subsection.

4.3.1 Input data

Recall that, for each commodity $c \in C$, the demand d_c is known, and $\mathcal{P}_c$ is the set of paths to reach D_c from O_c. For each path $p \in \mathcal{P}_c$, the *generalised cost* of travelling by p is given by τ_{cp}, and $\mathcal{P}_{c,a}$ denotes the set of paths using arc a, $a \in A$. A *train line* is a sequence of stations served by a same train. Let L be an index set representing the set of all train lines. We assume that a train should provide service to the entire line. We further define for each arc $a \in A$, the set of train lines using a, noted L_a, (with $L_a \in L$). For the sake of simplicity, this preliminary study assumes that the air and rail fleets are composed of exactly one aircraft type and one train type. For any arc $a \in A^{\text{air}} \cup A^{\text{rail}}$, the capacity of

a vehicle (aircraft of train) is known and denoted by κ_a and the cost c_{ai} is given for each criterion $i \in I$.

In order to take into account operational constraints, air and rail frequencies are limited. For instance, headway separation between two consecutive trains and the number of tracks limit the number of rail trips that can be scheduled on each track. Similarly, for flights, minimum separation constraints or airport and sector maximum capacities limit the number of flights that can be scheduled each day. As the frequency planning is done several months before operations, we do not model this level of detail in our constraints. However, we limit the flight frequency on each route in order to ensure that a feasible solution can be built. In the following, f^{air} and f^{rail} denote the maximum flight and train frequencies, per flight route and rail track, to take into account capacity constraints. In addition, F and R denote the maximum number of flights and trains that can be scheduled for the whole day, respectively.

4.3.2 Decision variables

For each commodity $c \in C$, for each path $p \in \mathcal{P}_c$, we define a continuous decision variable x_{cp} that corresponds to the share of passengers of commodity c assigned to path p. We then define for each arc $a \in A$, an auxiliary continuous decision variables v_a, giving the number of passengers assigned to travel through arc a. We also define an integer decision variable, y_a, that counts the number of services (flight or train frequency) to schedule on arc $a \in A$. Finally, for each train line $l \in L$, we define an integer decision variable f_l that assigns the frequency on l.

4.3.3 Objective function and constraints

The aim is to route passengers on the network according to their preferences (represented by costs) while minimising the CO_2 emissions of the transportation network. The problem is formulated as a bi-criterion optimisation problem where the criteria to be minimised are the total passenger generalised cost :

$$F_1(x, v, y, f) = \sum_{c \in C} d_c \sum_{p \in \mathcal{P}_c} x_{cp} \tau_{cp}, \tag{4.3}$$

and the global CO_2 emission :

$$F_2(x, v, y, f) = \sum_{a \in A} y_a c_{a,\text{CO}_2}. \tag{4.4}$$

We propose the following Mixed-Integer Linear Programming (MILP) formulation :

$$\min_{x,v,y,f} (F_1, F_2) \tag{4.5}$$

subject to :

$$\sum_{p \in \mathcal{P}_c} x_{cp} = 1 \qquad c \in C, \tag{4.5a}$$

$$\sum_{c \in C} \sum_{p \in \mathcal{P}_{c,a}} x_{cp} d_c = v_a \qquad a \in A, \tag{4.5b}$$

$$v_a \leq \kappa_a y_a \qquad a \in A, \tag{4.5c}$$

$$\sum_{a \in A^{\text{air}}} y_a \leq F, \tag{4.5d}$$

$$\sum_{l \in L} f_l \leq R, \tag{4.5e}$$

$$y_a = \sum_{l \in L_a} f_l \qquad a \in A^{\text{rail}}, \tag{4.5f}$$

$$0 \leq x_{cp} \leq 1 \qquad c \in C, p \in \mathcal{P}_c, \tag{4.5g}$$

$$y_a \in \left\{0, 1, \ldots, f^{\text{air}}\right\} \qquad a \in A^{\text{air}}, \tag{4.5h}$$

$$y_a \in \left\{0, 1, \ldots, f^{\text{rail}}\right\} \qquad a \in A^{\text{rail}}, \tag{4.5i}$$

$$f_l \in \left\{1, \ldots, f^{\text{rail}}\right\} \qquad l \in L. \tag{4.5j}$$

Constraints (4.5a) implement the definition of share of passengers and ensure that all passengers are routed in the network. Constraints (4.5b) link the share of passengers routed on each path with the total number of passenger travelling through arc a. Constraints (4.5c) implement the capacity bound on each arc. Constraints (4.5d) and (4.5e) ensure that the total number of flights scheduled and the total number of trains scheduled do not exceed the maximal total flight frequency and maximal total train frequency, respectively. Constraint (4.5f) ensure that a train is scheduled for an entire line. Finally, constraints (4.5g)-(4.5j) specify the definition domain of the decision variables.

4.3.4 Resolution approach

To obtain a realistic solution and reduce the computation time, we propose the following pre-processing. For each commodity c, there are generally several possible paths to travel from O_c to D_c. However, one can assume that passengers prefer to travel through shortest paths, according to the passenger generalised cost defined above. Therefore, for each commodity $c \in C$, a subset of k shortest paths from O_c to D_c is computed, where the shortest path is defined in terms of lowest generalised cost (equation 4.2). This computation is made using Yen algorithm (Yen, 1971), and the number k, of shortest paths computed is set by the user.

The bi-criterion optimisation problem introduced in Subsection 4.3.3 is addressed via

a weighted sum of the two criteria and reads as follows :

$$\min_{x,v,y,f} \lambda F_1 + (1-\lambda) F_2 \tag{4.6}$$

where λ is a user-defined parameter. The resolution of the optimisation problem is made using the MILP solver Gurobi (Gurobi Optimization, LLC, 2023), version 9.1.2. The global resolution framework is presented in Figure 4.2. Note that if λ parameter is set to 1, the

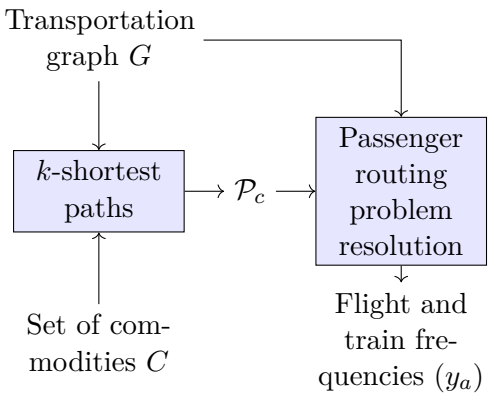


Figure 4.2 – Resolution framework of the ARSND problem.

CO_2 global criterion is not considered in the objective function. In this case, there may be multiple solutions minimising the total passenger cost, some of which are undesirable. The solver may plan more trains than necessary to accommodate all passengers since the CO_2 cost of these additional trains is not counted. To palliate this problem, we propose the following simple post-processing : launch the solver to minimise the total frequency, constraining the value of F_1 to be as good as previously found.

4.4 Spanish transportation network case study

The methodology is tested on the case study of the Spanish long-distance transportation network. Data and hypotheses are detailed, followed by a description and an analysis of the results obtained.

4.4.1 Spanish transportation network

Train and flight schedule data is collected for January 20, 2023 from the website of RENFE (RENFE, 2023a), the most important rail operator in Spain, and from OAG (OAG, 2023), respectively. Processing flight data reveals that the most used aircraft type for that day is a Boeing 737-800, which is used for 31% of the flights. The aircraft characteristics

(Boeing, 2023) of the B737-800 are therefore used to model CO_2 emission. Regarding rail CO_2 emission, rail services are operated by electric trains in Spain (RENFE, 2023b). The value of 358g CO_2/km for an electric train, computed by the Office of Rail and Road (Office of Rail and Road, 2022), is used for the study. For that day, 1,033 flights and 1,584 long-distance train trips were scheduled in Spain. These values are used for the daily maximum number, F, of flights, and the daily maximum number, R, of train trips, that can be scheduled. The train capacity on each arc $a \in A^{\text{rail}}$ is set to $\kappa_a = 500$ passengers, and the average flight capacity on each flight route $a \in A^{\text{air}}$ to $\kappa_a = 189$ passengers, as this is the value found from OAG for the Boeing 737-800 on that day. A load factor of 100% is assumed on each mode.

For simplicity, this study restricts the node set to 48 cities of Spain, that include the largest ones and at least one city per island. These cities are represented on Figure 4.3. The values of the parameters set for the study are defined according to the initial schedules and are presented in Table 4.1.

Table 4.1 – Parameters of the case study : average transfer and processing times at stations (minutes), maximum frequencies, and the number of shortest paths k computed. A distinction is made between the air transportation network : airports and flight routes (top line), and the rail transportation network : train stations and rail tracks (bottom line).

	t^{dep}	t^{arr}	t^{transfer}	Maximum route frequency	Maximum total frequency	k
Air network	90	20	60	25	1033	20
Rail network	10	0	15	50	1584	

4.4.2 Travel demand from mobile phone data

Passenger demand flows are measured by analysing anonymised Mobile Network Data (MND) collected by one of the main Mobile Network Operators (MNOs) in Spain. The MND consists of all the interactions between mobile devices and the antennas of the MNOs. The data is analysed using a processing pipeline for reconstructing door-to-door passenger journeys as in Burrieza-Galán *et al.* (2022). This method analyses the sequences of mobile phone records generated by anonymous mobile device users over a large period of time (several weeks) to infer their home location. It focuses on the activities and trips that can be detected from the daily sequences of records, in order to determine users' mobility patterns during a defined study period. The resulting activity-trip diaries are expanded to the total population based on home location, by comparing the available sample of residents in each census unit with the population figures. These diaries are aggregated in space and time to produce daily or hourly trip counts between the defined zones. For this study, data

from January 26, 2022 are collected. Schedules of January 2022 were not available, but it is reasonable to assume that schedules of January 2023 are sensibly the same as those of January 2022, due to the seasonality of schedules. The pipeline is configured to retrieve OD matrices at the district level between the 48 cities selected, covering only trips longer than 50 km, given the focus of this study on long-distance travel. The OD matrices include a segmentation by transport modes (road, rail, air, and rail-air multimodal trips), derived from map-matching techniques that compare the sequence of mobile phone records with the supply of each mode (*e.g.*, network, schedules, etc.). The OD matrices also include segmentation by travel time, using 1-hour windows. Figure 4.3 displays the travel demand between the 48 cities. In total, more than 97,000 people travel on that day, with a majority

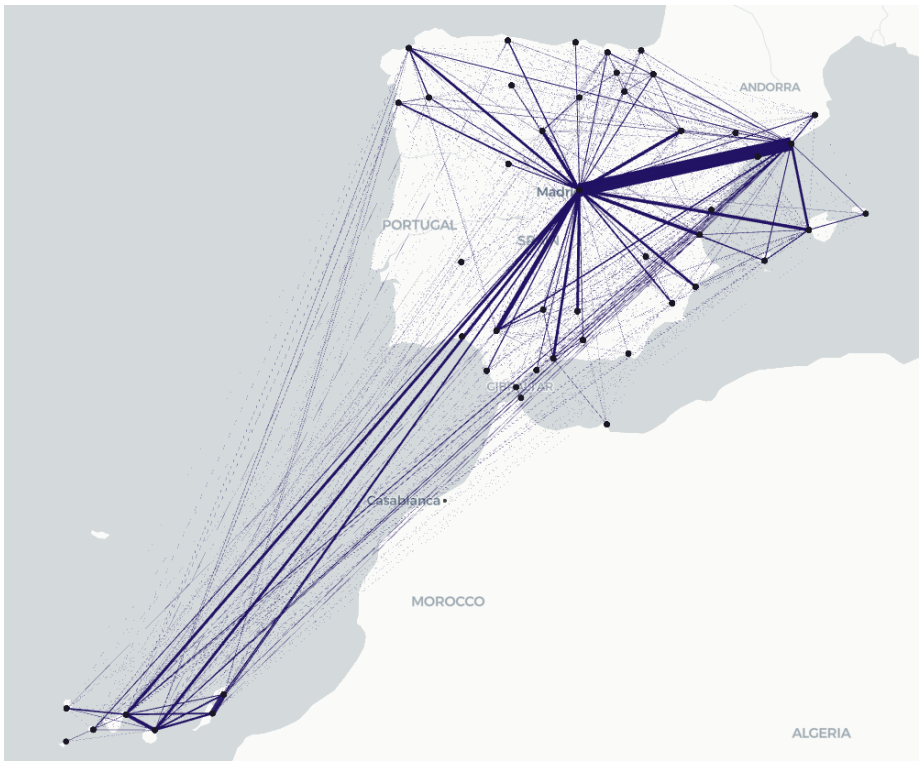


Figure 4.3 – OD matrix demand and the 48 cities considered. Bolder arcs correspond to higher number of passengers.

of trips between Madrid and Barcelona. Less than 1% of trips combines air and rail modes, and most of air-rail transfers take place in Madrid, as presented in Table 4.2. In order to consider these transfers, associated train stations are duplicated in the transportation network graph model. This demand serves as an input for the optimisation problem.

Table 4.2 – Multimodal passengers volume per city.

City	Number of transferring passengers
Madrid	453
Barcelona	92
Santiago de Compostela	56
Sevilla	45
Málaga	23
Bilbao	12
Zaragoza	6
Murcia	5
Valencia	4
Alicante	4

4.4.3 Optimisation results

Computations are performed on a laptop equipped with an AMD Ryzen 5 4500U CPU and 16 GB RAM. In a first step, only travel time is considered in the passengers' generalised cost function ($I = \{\text{Travel time}\}$). The time limit is set to two hours, and the gap limit is set to 0.5%. Computation information for several values of λ are summarised in Table 4.3. As the value of parameter λ increases (*i.e.* the weight of the travel time increases),

Table 4.3 – CPU times and optimality gaps according to the weighting parameter ($\lambda = 0$ corresponds to CO_2 emissions criterion, no consideration of the passenger cost criterion).

λ	Time (s)	MIP gap(%)
0	7200.0	1.03
0.1	7200.0	0.95
0.2	7200.0	0.77
0.3	7200.0	0.68
0.4	7200.0	0.53
0.5	7200.0	0.59
0.6	5923.8	0.50
0.7	585.7	0.49
0.8	94.8	0.34
0.9	10.8	0.33
1.0	0.4	0.00

the computation time decreases. Indeed, a further analysis reveals that the solver assigns each commodity to its shortest path. This solution is feasible since the capacity of the

transportation network is higher than required, showing thereby that the demand might have been underestimated. Conversely, as the value of parameter λ decreases, the solver tends to reach the computation time limit. Indeed, the shortest path generally uses a flight as the main transportation mode. Therefore, the solver searches among all feasible paths of each commodity to obtain the solution with the lowest CO_2 emission, increasing thereby the computation time.

4.4.4 Pool of solutions

Figure 4.4 displays the values of passenger and CO_2 emission criteria as a function of the weighting parameter λ (as the value of parameter λ decreases, the weight of the environmental cost in the objective function increases (Equation 4.6)). The sensitivity analysis shows that slightly considering the environmental cost ($\lambda = 0.9$) has a limited impact on passengers travel time for a significant saving of CO_2 emission. According to the model, for an increase in total passenger door-to-door travel time of around 200 hours, more than 500 tonnes of CO_2 can be saved. This is equivalent to an increase of 20 minutes

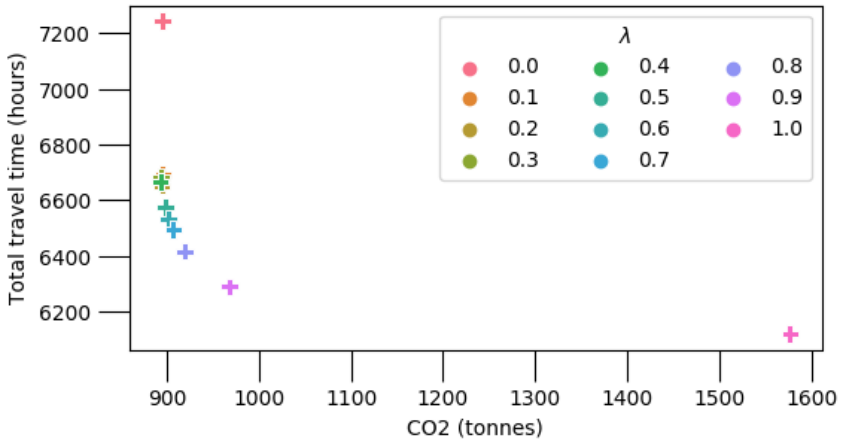


Figure 4.4 – Objective function criteria as a function of λ parameter.

per passenger of the average door-to-door travel time (215 minutes on average with $\lambda = 1$, for 235 minutes on average for $\lambda = 0.9$). Note that, road paths are not considered here. In practice, the door-to-door travel time by road could be lower and passengers may choose to travel by road when possible. The total travel time criterion would then be reduced, but remark that the CO_2 cost for travelling by road (which is not taken into consideration here) is not equal to zero.

In the following computational results, we therefore choose to set the λ parameter to 0.9.

4.4.5 Passenger trips

In the remaining of the study, results are compared with the initial supply and demand data. Table 4.4 summarises the number of passengers per mode. In the initial planning, only

Table 4.4 – Passenger volume per travel mode.

	Initial schedule	Optimised schedule
Air only	46,898	40,059
Rail only	50,148	45,357
Air-Rail	700	12,329

700 passengers use a combination of air and rail to travel. In the optimised planning, this number raises to more than 12,000 passengers. Consequently, the number of passengers using only rail on their journey is reduced by 10%. The number of trips using flights exclusively is reduced by 13%. Rail trips are less impacted by the solution as the CO_2 cost for travelling by air is higher than the one by train. Note that some journeys will not be affected by the synchronisation. In fact, some trips can only be made with one mean of transportation. This is particularly true for trips to/from islands, which can only be reached by air (boats are not considered here). Similarly, nearby cities are connected by train, and scheduling a flight on these routes is not relevant. In total, 57% of the passengers are not affected by the synchronisation.

Figure 4.5 presents the distribution of the total travel times for both the initial and the optimised plannings. Note that for the initial schedule, travel times are obtained with mobile phone data. For January 26, 2022, most of trips above 50 km using public transportation last between four and six hours. This observation still stands with the optimised air-rail frequency planning. For passengers using only one leg on their journey, the average door-to-door travel time is 190 minutes, compared with 350 minutes for passengers with at least two legs. In addition, only 856 passengers have a door-to-door travel time above eight hours in the optimised schedule compared with 8,500 in the initial one. One therefore observes that multimodal solutions result in a significant reduction in CO_2 emissions with minimal impact on passengers' door-to-door travel times, leading to an increase in multimodal demand, as shown in Table 4.4.

4.4.6 Integrated transportation network

Figure 4.6 displays the difference between the flight frequency before and after optimisation for each OD pair. Note that this schedule depends on the actual demand for that specific day. The results show a significant reduction in the number of daily flights between Barcelona (BCN) and Madrid (MAD). Actually, there are 30 fewer flights on this OD segment (considering both directions) after optimisation. By minimising travel times,

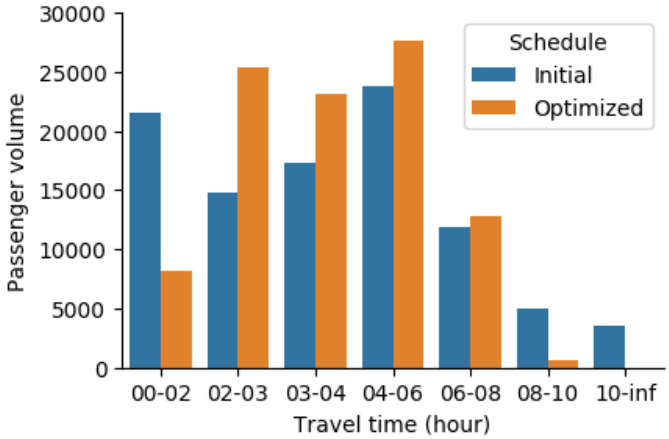


Figure 4.5 – Passenger volume per travel time window.

all passengers are routed by train, with an average travel time of 170 minutes, compared with 205 minutes by air, and producing 10 times less CO_2. This observed phenomenon is due to the inclusion of station processing time in the optimisation process. This highlights that the door-to-door journey time can be significantly different from the in-vehicle journey time.

One can also observe a large reduction in the number of flights between Spanish islands airports (LPA, ACE, IBZ, PMI, TNF, FUE). Figure 4.3 reveals that a small number of passengers travelled between Spanish islands on that day, probably due to the winter season. In addition, direct flights from mainland to islands are scheduled only from a small number of cities in the optimised schedule. In particular, Figure 4.7 displays the scheduled train and flight frequencies before and after optimisation. In the optimised planning, flights between mainland and the Canary islands are departing from Madrid, and flights to the Balearic islands are departing from Barcelona. As the CO_2 model favours short distance flights when no train options are available, travelling from Madrid to the Balearic islands costs less CO_2 if passengers use a train from Madrid to Barcelona first, then catch a flight to Balearic islands. The same phenomenon is observed in the opposite direction for the Canary islands.

Table 4.5 – Total flight and train frequencies.

	Initial schedule	Optimised schedule
Flights	1033	498
Trains	1584	619

The total train and flight frequencies for the initial and optimised planning are sum-

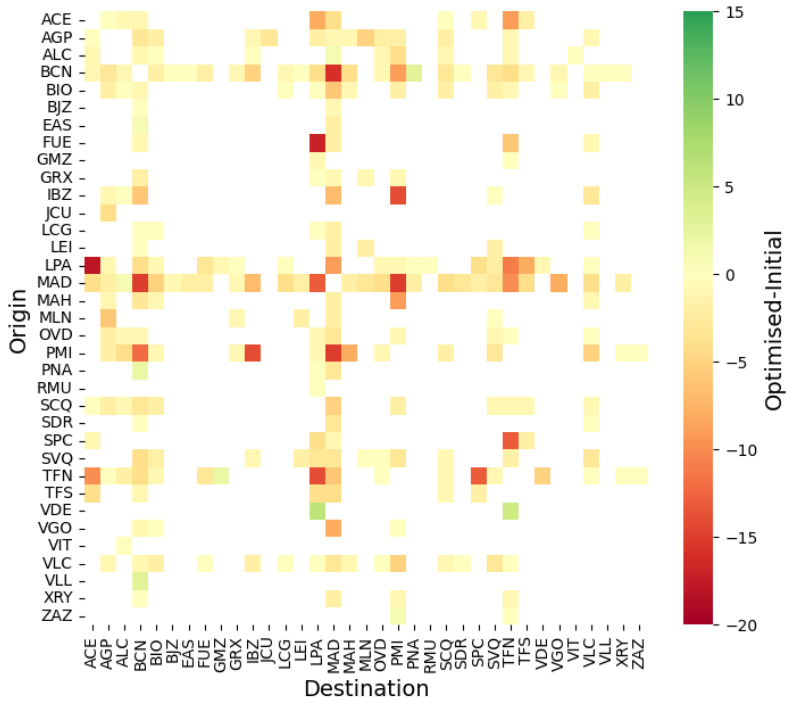
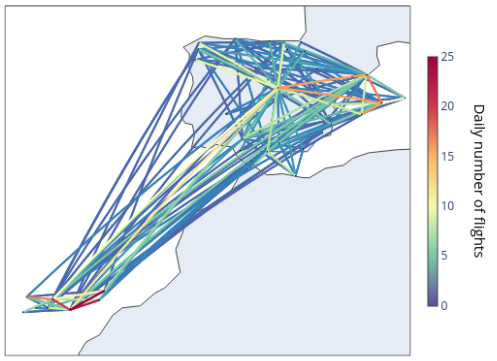


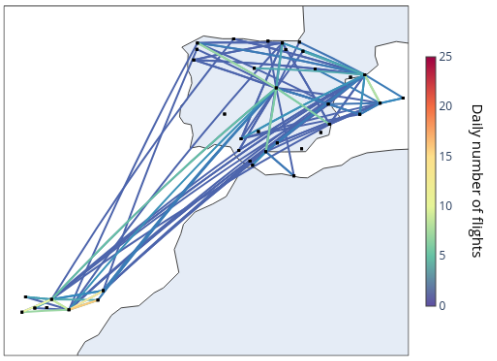
Figure 4.6 – Reduction in the number of daily flights per OD pair.

marised in Table 4.5. Both the total number of scheduled flights and trains are reduced in the optimised planning, compared with the initial schedule of January 2023. Regarding flights, the reduction can be explained by the fact that, as CO_2 is taken into account in the objective function, the algorithm reduces the number of flights whenever possible. The total train frequency is also reduced compared with the initial schedule. One of the main reasons is, as only trips among the 48 largest cities are studied, short distance trips are left out. For instance, in the initial schedule, 57 trains are scheduled between Segovia and Madrid. However, Segovia is not among the 48 largest cities of Spain ; Thus, passengers travelling to/from Segovia are therefore accounted in our data as passengers travelling to/from Madrid area. Therefore, these originally planned trains are no longer included in the new schedule. By considering the same CO_2 model as an estimator for the transportation network of January 2023, the CO_2 cost reduction is evaluated to 1,800 tonnes. As explained earlier, this value is probably overestimated as the volume of travelling passengers might be higher. However, results from the previous section reveals that CO_2 savings can be made by replacing short-haul flights by trains, with a limited impact on passenger door-to-door journey duration, as demonstrated on the Madrid-Barcelona segment.

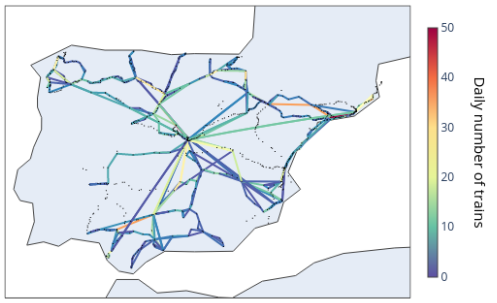
Figure 4.8 displays the volume of trips starting within each region of Spain. One can observe that multimodality (a transfer between a train and a flight) occurs most of the time



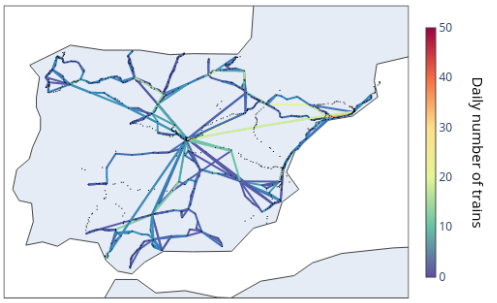
(a) Flight frequencies - Initial schedule



(b) Flight frequencies - Integrated schedule



(c) Train frequencies - Initial schedule



(d) Train frequencies - Integrated schedule

Figure 4.7 – Flights (top line) and train (bottom line) frequencies of the initial (left column) and optimised (right column) plannings.

in Barcelona, while before optimisation the multimodality was mainly represented in the Madrid area. This figure corroborates the results obtained. Indeed, passengers who want to travel to the Balearic islands take a flight from Barcelona in the optimised schedule. However, as no more flights are scheduled between Barcelona and Madrid in the integrated planning, passengers arriving from the Madrid area must first take a train to reach Barcelona, then connect with a flight.

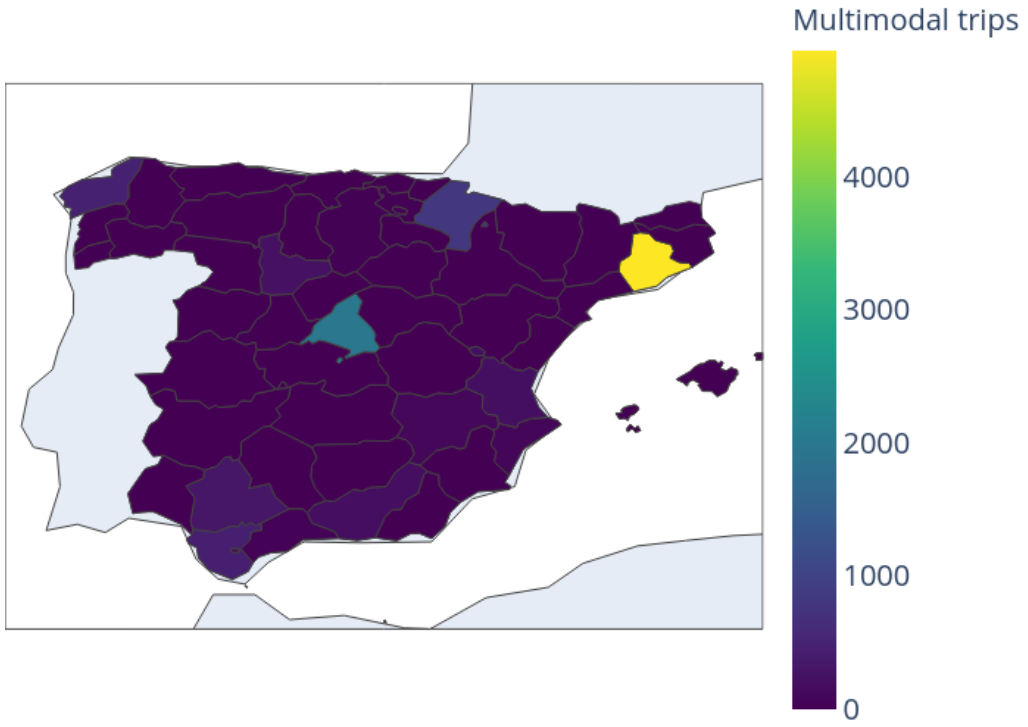


Figure 4.8 – Volume of multimodal trips starting within each NUTS-3 (Nomenclature of Territorial Units for Statistics-Level 3) region after optimisation.

4.5 Conclusion

In this chapter, we introduced a mixed-integer linear formulation of the air-rail integrated frequency planning problem that considers both the passenger perspective and CO_2 emissions. The frequency planning model proposed is able to consider a large range of possible passenger preferences in terms of travel time, price and environmental awareness. Furthermore, travel time is estimated door-to-door, and not solely from one station to another, including thereby potential transfers and station processing times. The model is implemented and tested on the case study of the Spanish transportation network. Insights on passenger demand are obtained through the analysis of mobile phone data, used as input data for the optimisation model. Results show that considering CO_2 emission while designing long-distance schedules succeeds in reducing by several tonnes the carbon footprint of the transportation system, at the expense of a moderate increase of the average door-to-door travel time of passengers : only 20 minutes per passenger on average. In particular, short-distance flights such as Barcelona-Madrid are no longer planned, as there is a relevant alternative by train. In addition, the number of trips combining air and rail is

increased.

Numerous future tracks of research are envisaged. First, note that the scope of the study is limited to Spain due to data availability. If an integrated air-rail network should be developed at the European scale, large-scale demand data is required. Second, the study is limited to air and rail modes, but one can easily include other transportation means in the model, such as road. Then, we propose a simple CO_2 emission model as a first estimator for this preliminary study but it could be improved by including fleet scheduling issues in the optimisation process. Furthermore, the global warming potential could be an alternative metric to capture better the environmental effect of transportation, including non-CO_2 effects. In addition, this chapter only considers CO_2 emissions in short-term planning, but a full life cycle analysis should be further developed, especially in the case of infrastructure construction, as the long-term impact of a new line can be environmentally profitable. Finally, the main future work foreseen is the refinement of the frequency calculation per time window. While our current analysis focuses on the daily frequency, there is still uncertainty regarding the synchronisation of air and rail services. Therefore, a future step is to divide the day into time window intervals and calculate the required air and rail frequencies to satisfy the demand. In addition, the expected arrival time preferences of passengers can also be included in the model. The output of such a model will also facilitate the timetabling process, as an hourly frequency will already be determined.

Conclusion

Outcomes and discussion

Airport capacity issues, growing environmental awareness, and the COVID-19 outbreak are leading to a rethink of the air transportation system, and in particular, its relationship with rail. Collaboration is replacing competition, and incentives to develop coordination mechanisms have been introduced. The objective of this dissertation was therefore to introduce mathematical models and to develop synchronisation tools between air and rail, to the benefit of passengers.

First of all, we have highlighted several challenges that operators may face on their way towards a strong collaboration, especially regarding data availability. Indeed, Chapter 1 showed that the timetable generation process relies on accurate data, both from the supply and the demand sides. The availability of reliable and accurate data is a basis for the further development of coordination mechanisms.

Chapter 2 emphasised the difficulty of finding recent and accurate supply data at a limited monetary cost. In addition, in Europe, the multimodal demand between air and rail remains today largely unknown. To address these issues, we developed several simulation techniques to simulate realistically missing information such as the ID of vehicles operating flights and trains, or the multimodal passenger demand. We made publicly available the data set obtained for the operations research community to contribute on these challenging issues, and to allow transportation planners to test further air-rail synchronisation mechanisms on realistic instances.

In Chapter 3, we proposed two models for timetable synchronisation, with the constraint of limited changes to existing train and flight schedules. The first introduced mechanism, suitable for long-term or medium-term planning, aims at improving the quality of a passenger connection, based on the available connection time. The second proposed mechanism is a tactical tool : it aims at determining in real time the new departure time of trains and flights so as to wait for delayed passengers, and to reduce the overall passenger delay. These two mechanisms take into account operational constraints such as : airport and train station capacities, vehicle turnaround times, or the guarantee of intramodal passenger connections. We tested our methodologies on the Western Europe case study, covering three countries, more than 10,000 flights and 500 trains, 496 airports and 72 train stations.

The results showed that large benefit can already be achieved for passengers at limited costs for transportation service providers.

Finally, in Chapter 4, we went further in air-rail synchronisation by solving the line planning problem of airline and railway operators in a joint way. We proposed to minimise a novel metric, the generalised passenger travel cost, which can include a broad spectrum of passengers' preferences in terms of door-to-door travel time, monetary cost, number of transfers, CO_2 awareness, etc. In addition, we took into account the total CO_2 cost in the objective function. Results highlighted that considering the passenger door-to-door travel time, and not only the in-vehicle time, can significantly change the transportation network. Short-haul flights are replaced by trains, and flights are maintained where no feasible (overseas) or reasonable (in terms of door-to-door travel time) train service can be offered to passengers.

Discussion and future work

This dissertation highlighted several challenges that need to be addressed in order to efficiently synchronise air and rail schedules. First, the need for up-to-date (near real-time) data on passenger demand is crucial. The COVID-19 crisis shows the change in passenger travel patterns : the hypothesis stating that the demand steadily increases has been invalidated. An accurate estimate of the demand can allow transportation providers to adapt their services. Second, as shown in Chapter 2, a standardised repository of transportation data, from several operators and countries, will greatly facilitate the implementation of coordination mechanisms. Finally, the size of the optimisation problems to be addressed requires resolution algorithms able to handle such a large amount of data. The results of the dissertation highlighted the difficulty of classical techniques to scale up to real-life instances and the implementation of an adaptive system implies scheduling operations in a limited amount of time (less than a day).

Beyond this dissertation, there are several avenues for future work.

First, the application of robust optimisation to some of the problems addressed may be relevant. Indeed, the results obtained in this thesis are intrinsically linked to the input data. Ben-Tal *et al.* (2009) insist on how a small change in the data can lead to significantly different solutions. A sensitivity analysis should be carried out to analyse its impact of changes in the data on the solutions found, so as to propose a more robust solution in the case of anticipated variations of the demand. In addition, our models rely on train and flight schedules, but observed delays have not been taken into account. For example, Section 3.1 of Chapter 3 suggests making small schedule adjustments to improve multimodal passenger connections at a strategic level. However, including historical delay patterns in the optimisation process can improve the robustness of the schedule. Therefore, investigating the benefits of robust optimisation could be relevant to address the Air-Rail Strategic Timetable Synchronisation (ARSTS) problem. Similarly, including observed Air Traffic

Flow Management (ATFM) delay in the optimisation problem of Section 3.2 of Chapter 3 will improve the quality of the tool, and enhance its potential for practical implementation in real-world scenarios.

Second, we focused mainly on the airline and air-passenger perspectives, although we tried to include rail constraints as much as possible. However, rail-rail connections are obviously also of primary interest for railway operators. Our models and methods could therefore be improved by taking also into consideration rail passenger connections at major train stations. In addition, some operational constraint such as the minimum headway constraint on tracks could be included in our optimisation models.

Finally, Chapter 4 is based on several simplifying assumptions, such as the average connection time between trains and flights, constant passenger travel time and processing time at stations, unique type of aircraft, etc. These assumptions need to be gradually removed in favour of more realistic estimates. A possible solution to address this issue is to refine the model per time-window interval, *i.e.,* to calculate the train and flight frequencies per time window within a day. By reducing the time-window duration, the output of such a model could then also lead to a final integrated air-rail timetable.

Bibliographie

Adrangi, B., Chatrath, A., and Raffiee, K. (2001). The demand for US air transport service : A chaos and nonlinearity investigation. *Transportation Research Part E : Logistics and Transportation Review*, 37(5) :337–353.

AECFA (2023). Slot coordination-Spanish airports. `https://www.slotcoordination.es/en/slot/library`.

Alexander, L., Jiang, S., Murga, M., and González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C : Emerging technologies*, 58 :240–250.

Aliari, Y. and Haghani, A. (2012). Bluetooth sensor data and ground truth testing of reported travel times. *Transportation Research Record*, 2308(1) :167–172.

Allard, R. F. and Moura, F. M. M. V. e. (2014). Optimizing high-speed rail and air transport intermodal passenger network design. *Transportation Research Record*, 2448(1) :11–20.

Alumur, S. and Kara, B. Y. (2008). Network hub location problems : The state of the art. *European Journal of Operational Research*, 190(1) :1–21.

Andersen, J., Crainic, T. G., and Christiansen, M. (2009). Service network design with management and coordination of multiple fleets. *European Journal of Operational Research*, 193(2) :377–389.

Ashiabor, S., Baik, H., and Trani, A. (2007). Logit models for forecasting nationwide intercity travel demand in the United States. *Transportation Research Record*, 2007(1) :1–12.

Aston, J. A. and Koopman, S. J. (2006). A non-Gaussian generalization of the airline model for robust seasonal adjustment. *Journal of Forecasting*, 25(5) :325–349.

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A., and Zou, B. (2010). Total delay impact study : A comprehensive assessment

of the costs and impacts of flight delay in the United States. Technical report, University of California, Berkeley. Institute of Transportation Studies.

Banerjee, N., Morton, A., and Akartunalı, K. (2020). Passenger demand forecasting in scheduled transportation. *European Journal of Operational Research*, 286(3) :797–810.

Banister, D. and Berechman, J. (2003). *Transport investment and economic development*. Routledge.

Barcelö, J., Montero, L., Marqués, L., and Carmona, C. (2010). Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. *Transportation Research Record*, 2175(1) :19–27.

Behrens, C. and Pels, E. (2012). Intermodal competition in the London–Paris passenger market : High-Speed Rail and air transport. *Journal of Urban Economics*, 71(3) :278–288.

Belobaba, P., Odoni, A., and Barnhart, C. (2009). *The global airline industry*. John Wiley & Sons.

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*. Princeton University Press.

Boeing (2023). Next-generation 737 airplane characteristics for airport planning. `https://www.boeing.com/resources/boeingdotcom/commercial/airports/acaps/737NG_REVA.pdf`.

Boussemart, F., Hemery, F., Lecoutre, C., and Sais, L. (2004). Boosting systematic search by weighting constraints. In *European Conference on Artificial Intelligence*, volume 16, pages 146–150.

Brännlund, U., Lindberg, P. O., Nou, A., and Nilsson, J.-E. (1998). Railway timetabling using Lagrangian relaxation. *Transportation Science*, 32(4) :358–369.

Budde, A., De Wit, J., and Burghouwt, G. (2008). Borrowing from behavioural science : A novel method for the analysis of indirect temporal connectivity at airport hubs. In *Air Transport Research Society Conference*, Athens, Greece.

Bueno, J., Burrieza, J., García Cantú, O., Livingston, C., Balac, M., Scozzaro, G., and Buire, C. (2022). SESAR TRANSIT Final Project Results Report. `https://www.transit-h2020.eu/deliverables`.

Buire, C. (2023). Air-Rail Passenger Demand Instances. `https://doi.org/10.57745/5WB9KG`.

Buire, C., Delahaye, D., Marzuoli, A., Feron, E., and Mongeau, M. (2022). Air-rail timetable synchronization for a seamless passenger journey. In *Proceedings of International Workshop on ATM/CNS*, pages 79–86, Tokyo, Japan.

Buire, C., Delahaye, D., Mongeau, M., Marzuoli, A., Bueno-Gonzalez, J., Artime, R., G. Cantú Ros, O., and Burrieza-Galán, J. (2023). Leveraging passengers' mobile network data for an integrated air-rail frequency planning in Spain. In 13^{th} *SESAR Innovation Days*, Sevilla, Spain.

Buire, C., Marzuoli, A., Delahaye, D., and Mongeau, M. (2024). Air–rail timetable synchronisation : Improving passenger connections in Europe within and across transportation modes. *Journal of Air Transport Management*, 115 :102526.

Buire, C., Scozzaro, G., Marzuoli, A., Feron, E., and Delahaye, D. (2021). A year into the pandemic : A passenger perspective on its impact at Paris-Charles de Gaulle airport. In *2021 IEEE International Conference on Big Data*, pages 2925–2935. IEEE.

Burghouwt, G. and de Wit, J. (2005). Temporal configurations of European airline networks. *Journal of Air Transport Management*, 11(3) :185–198.

Burghouwt, G. and Redondi, R. (2013). Connectivity in air transport networks : An assessment of models and applications. *Journal of Transport Economics and Policy*, 47(1) :35–53.

Burrieza-Galán, J., Jordá, R., Gregg, A., Ruiz, P., Rodríguez, R., Sala, M., Torres, J., García-Albertos, P., Ros, O. C., and Herranz, R. (2022). A methodology for understanding passenger flows combining mobile phone records and airport surveys : Application to Madrid-Barajas airport after the COVID-19 outbreak. *Journal of Air Transport Management*, 100 :102163.

Bussieck, M. R., Kreuzer, P., and Zimmermann, U. T. (1997). Optimal lines for railway systems. *European Journal of Operational Research*, 96(1) :54–63.

Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, 72(2) :387–405.

Cao, Z., Ceder, A., Li, D., and Zhang, S. (2019). Optimal synchronization and coordination of actual passenger-rail timetables. *Journal of Intelligent Transportation Systems*, 23(3) :231–249.

Caprara, A., Fischetti, M., and Toth, P. (2002). Modeling and solving the train timetabling problem. *Operations Research*, 50(5) :851–861.

Castelli, L., Pesenti, R., and Ukovich, W. (2004). Scheduling multimodal transportation systems. *European Journal of Operational Research*, 155(3) :603–615.

Ceder, A., Golany, B., and Tal, O. (2001). Creating bus timetables with maximal synchronization. *Transportation Research Part A : Policy and Practice*, 35(10) :913–928.

Ceder, A. and Wilson, N. H. (1986). Bus network design. *Transportation Research Part B : Methodological*, 20(4) :331–344.

Chankong, V. and Haimes, Y. Y. (1983). Optimization-based methods for multiobjective decision-making : An overview. *Large Scale Systems In Information And Decision Technologies*, 5(1) :1–33.

Chiambaretto, P., Baudelaire, C., and Lavril, T. (2013). Measuring the willingness-to-pay of air-rail intermodal passengers. *Journal of Air Transport Management*, 26 :50–54.

Cipriani, E., Crescenzi, L., and Nigro, M. (2014). Behavioral models for the estimation of the air transport demand : The case study of Rome-London flight connection. *Procedia-Social and Behavioral Sciences*, 111 :78–87.

Claessens, M., van Dijk, N., and Zwaneveld, P. (1998). Cost optimal allocation of rail passenger lines. *European Journal of Operational Research*, 110(3) :474–489.

Clewlow, R. R., Sussman, J. M., and Balakrishnan, H. (2014). The impact of high-speed rail and low-cost carriers on European air passenger traffic. *Transport Policy*, 33 :136–143.

COHOR - Association pour la coordination des horaires (2023). Aéroport Paris-Charles de Gaulle. `https://www.cohor.org/aeroport-paris-charles-de-gaulle-cdg/`.

Cook, A., Tanner, G., Cristóbal, S., and Zanin, M. (2012). Passenger-oriented enhanced metrics. In *Second SESAR Innovation Days*, Braunschweig, Germany.

Danesi, A. (2006). Measuring airline hub timetable co-ordination and connectivity : Definition of a new index and application to a sample of European hubs. Technical report, Edizioni Università di Trieste.

Data4PT (2023). Data models. `https://data4pt-project.eu/data-models/`.

Delgado, L., Gurtner, G., Mazzarisi, P., Zaoli, S., Valput, D., Cook, A., and Lillo, F. (2021). Network-wide assessment of ATM mechanisms using an agent-based model. *Journal of Air Transport Management*, 95 :102108.

Delgado, L., Martín, J., Blanch, A., and Cristóbal, S. (2016). Hub operations delay recovery based on cost optimisation-Dynamic cost indexing and waiting for passengers strategies. In *Sixth SESAR Innovation Days*, Delft, Netherlands.

Dennis, N. (1994). Airline hub operations in Europe. *Journal of Transport Geography*, 2(4) :219–233.

Desaulniers, G., Desrosiers, J., Dumas, Y., Solomon, M. M., and Soumis, F. (1997). Daily aircraft routing and scheduling. *Management Science*, 43(6) :841–855.

Dollevoet, T., Huisman, D., Kroon, L., Schmidt, M., and Schöbel, A. (2015). Delay management including capacities of stations. *Transportation Science*, 49(2) :185–203.

Dollevoet, T., Huisman, D., Schmidt, M., and Schöbel, A. (2012). Delay management with rerouting of passengers. *Transportation Science*, 46(1) :74–89.

Dray, L., Marzuoli, A., Evans, A., Laplace, I., and Feron, E. (2015). Air transportation and multimodal, collaborative decision making during adverse events. In *Eleventh USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)*, Lisbon, Portugal.

Dutta, G. and Ghosh, P. (2012). A passenger revenue management system (RMS) for a national railway in an amerging Asian economy. *Journal of Revenue and Pricing Management*, 11 :487–499.

Etschmaier, M. M. and Mathaisel, D. F. (1985). Airline scheduling : An overview. *Transportation Science*, 19(2) :127–138.

Eurocontrol (2023). Aviation Data for Research. `https://www.eurocontrol.int/dashboard/rnd-data-archive`.

European Commission (2004). Regulation (EC) No 261/2004 of the European Parliament and of the Council of 11 February 2004 establishing common rules on compensation and assistance to passengers in the event of denied boarding and of cancellation or long delay of flights, and repealing Regulation (EEC) No 295/91 (text with EEA relevance). `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32004R0261`.

European Commission (2023a). National Access Points. `https://transport.ec.europa.eu/system/files/2023-09/its-national-access-points-2023-09-19.pdf`.

European Commission (2023b). Single European Sky. `https://transport.ec.europa.eu/transport-modes/air/single-european-sky_en`.

European Commission, Directorate-General for Mobility and Transport (2020). Sustainable and smart mobility strategy – Putting European transport on track for the future. Technical report, European Commission.

Europe's Rail (2023). Webpage. `https://rail-research.europa.eu/`.

Eurostat (2019). Harmonised indices of consumer prices - inflation rate. `https://www.faa.gov/data_research/aviation_data_statistics/media/cost_delay_estimates.pdf`.

Flightradar24 (2023). Webpage. https://www.flightradar24.com.

FlightStats (2023). Webpage. https://www.flightstats.com/v2.

Fluko - Flughafenkoordination Deutschland GmbH (2023). Airport capacity parameters. https://fluko.org/en/flughaefen/flughafen-kapazitaets-parameter/.

Fraport (2021). Passengers at Frankfurt airport. https://www.fraport.com/en/newsroom/frafacts.html.

García, P., Herranz, R., and Javier, J. (2016). Big data analytics for a passenger-centric air traffic management system. *Sixth SESAR Innovation Days, Delft, Netherlands*.

Ghobrial, A., Balakrishnan, N., and Kanafani, A. (1992). A heuristic model for frequency planning and aircraft routing in small size airlines. *Transportation Planning and Technology*, 16(4) :235–249.

Ghoseiri, K., Szidarovszky, F., and Asgharpour, M. J. (2004). A multi-objective train scheduling model and solution. *Transportation Research Part B : Methodological*, 38(10) :927–952.

Givoni, M. and Banister, D. (2006). Airline and railway integration. *Transport Policy*, 13(5) :386–397.

Goossens, J.-W., van Hoesel, S., and Kroon, L. (2006). On solving multi-type railway line planning problems. *European Journal of Operational Research*, 168(2) :403–424.

Gopalan, R. and Talluri, K. T. (1998). Mathematical models in airline schedule planning : A survey. *Annals of Operations Research*, 76 :155–185.

Grosche, T. (2009). Airline scheduling process. In *Computational Intelligence in Integrated Airline Scheduling*, pages 7–46. Springer.

Groupe ADP (2021). Financial Release.

Gurobi Optimization, LLC (2023). Gurobi Optimizer Reference Manual.

Hane, C. A., Barnhart, C., Johnson, E. L., Marsten, R. E., Nemhauser, G. L., and Sigismondi, G. (1995). The fleet assignment problem : Solving a large-scale integer program. *Mathematical Programming*, 70(1) :211–232.

Hebrard, E., Hnich, B., O'Sullivan, B., and Walsh, T. (2005). Finding diverse and similar solutions in constraint programming. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI)*, volume 5, pages 372–377, Pittsburgh, Pennsylvania.

Huang, K., Wu, J., Liao, F., Sun, H., He, F., and Gao, Z. (2021). Incorporating multimodal coordination into timetabling optimization of the last trains in an urban railway network. *Transportation Research Part C : Emerging Technologies*, 124 :102889.

Huang, W., Lin, Y., Lin, B., and Zhao, L. (2019). Modeling and predicting the occupancy in a China hub airport terminal using Wi-Fi data. *Energy and Buildings*, 203 :109439.

Innaxis (2023). Kerb-to-gate travel time distribution. http://visual.innaxis.org/dataset2050/d2d-time-distribution/.

Interline (2022). Transitland. https://www.interline.io/transitland/.

International Air Transport Association (2020). Air passenger market analysis, december 2019.
https://www.iata.org/en/iata-repository/publications/economic-reports/air-passenger-monthly---dec-2019/.

International Air Transport Association (2022). IATA Carbon Offset Program.
https://www.iata.org/contentassets/922ebc4cbcd24c4d9fd55933e7070947/icop_faq_general-for-airline-participants.pdf.

International Air Transport Association (IATA) (2023a). Air passenger market analysis, August 2023.
https://www.iata.org/en/iata-repository/publications/economic-reports/air-passenger-market-analysis---august-2023/.

International Air Transport Association (IATA) (2023b). What is a slot ?
https://www.iata.org/en/programs/ops-infra/slots/whats-a-slot/.

International Civil Aviation Organization (2023). Présentation des résultats statistiques du transport aérien en 2019. https://www.icao.int/annual-report-2019/Documents/ARC_2019_Air%20Transport%20Statistics_fr.pdf.

International Energy Agency (IEA) (2020). Global energy-related CO_2 emissions by sector. https://www.iea.org/data-and-statistics/charts/global-energy-related-co2-emissions-by-sector.

International Energy Agency (IEA) (2023). Aviation.
https://www.iea.org/energy-system/transport/aviation.

Jansen, L. N., Pedersen, M. B., and Nielsen, O. A. (2002). Minimizing passenger transfer times in public transport timetables. In *7th Conference of the Hong Kong Society for Transportation Studies, Transportation in the information age, Hong Kong*, pages 229–239.

Kazda, A. and Caves, R. E. (2007). *Airport design and operation*, volume 2. Elsevier Amsterdam.

Ke, Y., Nie, L., Liebchen, C., Yuan, W., and Wu, X. (2020). Improving synchronization in an air and high-speed rail integration service via adjusting a rail timetable : A real-world case study in China. *Journal of Advanced Transportation*, 2020.

Keeney, R. L. (1973). A decision analysis with multiple objectives : The Mexico City airport. *The Bell Journal of Economics and Management Science*, pages 101–117.

König, E. (2020). A review on railway delay management. *Public Transport*, 12(2) :335–361.

Laplace, I., Marzuoli, A., and Féron, E. (2014). META-CDM : Multimodal, efficient transportation in airports and collaborative decision making. In *AUN 2014, Airports in Urban Networks*.

Lederer, P. J. and Nambimadom, R. S. (1998). Airline network design. *Operations Research*, 46(6) :785–804.

Leng, N., Nie, L., Guo, G., and Wu, X. (2015). Passenger flow forecasting for Chinese high-speed rail network. In *2015 International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC-15)*, pages 675–678, Shenyang, China. Atlantis Press.

Levin, A. (1969). Some fleet routing and scheduling problems for air transportation systems. Technical report, Massachusetts Institute of Technology, Flight Transportation Laboratory, Cambridge, MA.

Lewe, J.-h., Hivin, L., and Mavris, D. (2012). Multimodal transportation demand forecast using system dynamics and agent-based models. In *12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 5505, Indianapolis, IN.

Li, X., Jiang, C., Wang, K., and Ma, J. (2018). Determinants of partnership levels in air-rail cooperation. *Journal of Air Transport Management*, 71 :88–96.

Li, Z.-C. and Sheng, D. (2016). Forecasting passenger travel demand for air and high-speed rail integration service : A case study of Beijing-Guangzhou corridor, China. *Transportation Research Part A : Policy and Practice*, 94 :397–410.

Livingston, C., Balac, M., Gregg, A., Bueno, J., Burrieza, J., Scozzaro, G., and Buire, C. (2022). Impact assessment of new intermodal concepts and passenger information services : Conclusions and recommendations. Technical report, TRANSIT Consortium.

Lockheed, F. (2021). Unofficial Deutsche Bahn Fernverkehr (German railways long-distance trains) Timetable GTFS Feed. `https://github.com/fredlockheed/db-fv-gtfs`.

Lythgoe, W. and Wardman, M. (2002). Demand for rail travel to and from airports. *Transportation*, 29(2) :125–143.

Maertens, S., Grimme, W., and Bingemer, S. (2020). The development of transfer passenger volumes and shares at airport and world region levels. *Transportation Research Procedia*, 51 :171–178.

Magnanti, T. L. and Wong, R. T. (1984). Network design and transportation planning : Models and algorithms. *Transportation Science*, 18(1) :1–55.

Marzuoli, A., Boidot, E., Colomar, P., Guerpillon, M., Feron, E., Bayen, A., and Hansen, M. (2016). Improving disruption management with multimodal collaborative decision-making : A case study of the Asiana crash and lessons learned. *IEEE Transactions on Intelligent Transportation Systems*, 17(10) :2699–2717.

Marzuoli, A., Boidot, E., Feron, E., and Srivastava, A. (2019). Implementing and validating air passenger-centric metrics using mobile phone data. *Journal of Aerospace Information Systems*, 16(4) :132–147.

Marzuoli, A., Boidot, E., Feron, E., van Erp, P. B., Ucko, A., Bayen, A., and Hansen, M. (2015). Multimodal impact analysis of an airside catastrophic event : A case study of the Asiana crash. *IEEE Transactions on Intelligent Transportation Systems*, 17(2) :587–604.

Marzuoli, A., Monmousseau, P., and Feron, E. (2018). Passenger-centric metrics for air transportation leveraging mobile phone and Twitter data. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 588–595, Singapore, Singapore. IEEE.

Ministerio de Transportes, Movilidad y Agenda Urbana - NAP (2023). Datos de vuelos de AECFA. `https://nap.mitma.es/Files/Detail/920`.

MobilityData (2023). General Transit Feed Specification. `https://gtfs.org/`.

Monmousseau, P., Delahaye, D., Marzuoli, A., and Feron, E. (2019). Door-to-door travel time analysis from Paris to London and Amsterdam using Uber data. In *SID 2019, 9th SESAR Innovation Days*, Athens, Greece.

Monmousseau, P., Marzuoli, A., Feron, E., and Delahaye, D. (2020). Impact of COVID-19 on passengers and airlines from passenger measurements : Managing customer satisfaction while putting the US air transportation system to sleep. *Transportation Research Interdisciplinary Perspectives*, 7 :100179.

Montero, L., Ros-Roca, X., Herranz, R., and Barceló, J. (2019). Fusing mobile phone data with other data sources to generate input OD matrices for transport models. *Transportation Research Procedia*, 37 :417–424.

Montlaur, A. and Delgado, L. (2017). Flight and passenger delay assignment optimization strategies. *Transportation Research Part C : Emerging Technologies*, 81 :99–117.

Mota, M. M., Scala, P., Herranz, R., Schultz, M., and Jimenez, E. (2020). Creating the future airport passenger experience : IMHOTEP. In *European Modelling Simulation Symposium, Athens, Greece.*

Nachtigall, K. and Voget, S. (1996). A genetic algorithm approach to periodic railway synchronization. *Computers & Operations Research*, 23(5) :453–463.

Nadel, A. (2011). Generating diverse solutions in SAT. In *International Conference on Theory and Applications of Satisfiability Testing*, pages 287–301. Springer.

Nguyen, T. A., Do, M., Gerevini, A. E., Serina, I., Srivastava, B., and Kambhampati, S. (2012). Generating diverse plans to handle unknown and partially known user preferences. *Artificial Intelligence*, 190 :1–31.

Nikoue, H., Marzuoli, A., Clarke, J.-P., Feron, E., and Peters, J. (2015). Passenger flow predictions at Sydney international airport : A data-driven queuing approach. *arXiv preprint arXiv :1508.04839.*

OAG (2023). Schedule analyser. `https://www.oag.com/schedules-analyser`.

Odijk, M. A. (1996). A constraint generation algorithm for the construction of periodic railway timetables. *Transportation Research Part B : Methodological*, 30(6) :455–464.

Office of Rail and Road (2022). Rail emissions. `https://dataportal.orr.gov.uk/statistics/infrastructure-and-emissions/rail-emissions/`.

Okumura, M. and Tsukai, M. (2007). Air-rail inter-modal network design under hub capacity constraint. *Journal of the Eastern Asia Society for Transportation Studies*, 7 :180–194.

OpenSky Network, T. (2023). Webpage. `http://www.opensky-network.org`.

Park, Y. and Ha, H.-K. (2006). Analysis of the impact of high-speed railroad service on air transport demand. *Transportation Research Part E : Logistics and Transportation Review*, 42(2) :95–104.

Peeters, L. (2003). *Cyclic railway timetable optimization.* PhD thesis, Erasmus Research Institute of Management.

Pels, E., Nijkamp, P., and Rietveld, P. (2003). Access to and competition between airports : A case study for the San Francisco Bay area. *Transportation Research Part A : Policy and Practice*, 37(1) :71–83.

Phang, S.-Y. (2003). Strategic development of airport and rail infrastructure : The case of Singapore. *Transport Policy*, 10(1) :27–33.

Prud'homme, C. and Fages, J.-G. (2022). Choco-solver : A Java library for constraint programming. *Journal of Open Source Software*, 7(78) :4708.

Puettmann, C. and Stadtler, H. (2010). A collaborative planning approach for intermodal freight transportation. *OR Spectrum*, 32(3) :809–830.

RENFE (2022). Línea C-1 RENFE cercanías de Madrid. `https://www.redtransporte.com/madrid/cercanias-renfe/linea-c-1.html`.

RENFE (2023a). RENFE Data. `https://data.renfe.com/dataset?res_format=GTFS`.

RENFE (2023b). Train fleet. `https://www.renfe.com/es/en/renfe-group/renfe-group/fleet-of-trains`.

Rexing, B., Barnhart, C., Kniker, T., Jarrah, A., and Krishnamurthy, N. (2000). Airline fleet assignment with time windows. *Transportation Science*, 34(1) :1–20.

Román, C. and Martín, J. C. (2014). Integration of HSR and air transport : Understanding passengers' preferences. *Transportation Research Part E : Logistics and Transportation Review*, 71 :129–141.

Rossi, F., Van Beek, P., and Walsh, T. (2006). *Handbook of constraint programming*. Elsevier.

Sabre (2023). Webpage. `https://www.sabre.com/`.

Santos, B. F., Wormer, M. M., Achola, T. A., and Curran, R. (2017). Airline delay management problem with airport capacity constraints and priority decisions. *Journal of Air Transport Management*, 63 :34–44.

Schiewe, P. (2020). *Integrated optimization in public transport planning*, volume 160. Springer.

Schöbel, A. (2001). A model for the delay management problem based on mixed-integer-programming. *Electronic Notes in Theoretical Computer Science*, 50(1) :1–10.

Schöbel, A. (2009). Capacity constraints in delay management. *Public Transport*, 1 :135–154.

Scozzaro, G., Buire, C., Delahaye, D., and Marzuoli, A. (2023). Optimizing air-rail travel connections : A data-driven delay management strategy for seamless passenger journeys. In 13^{th} *SESAR Innovation Days*, Sevilla, Spain.

Scozzaro, G., Ma, J., Delahaye, D., Feron, E., and Mancel, C. (2022). Flight rescheduling to improve passenger journey during airport access mode disruptions. In *International Conference on Research in Air Transportation (ICRAT 2022)*, Tampa, FL.

Seaborn, C., Attanucci, J., and Wilson, N. H. (2009). Analyzing multimodal public transport journeys in London with smart card fare payment data. *Transportation Research Record*, 2121(1) :55–62.

Serafini, P. and Ukovich, W. (1989). A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4) :550–581.

SESAR Joint Undertaking (2020). MODUS-Modelling and assessing the role of air transport in an integrated, intermodal transport system. `https://modus-project.eu/`.

SESAR Joint Undertaking (2023a). MultiModX-Integrated passenger-centric planning of multimodal transport networks . `https://sesarju.eu/projects/MultiModX`.

SESAR Joint Undertaking (2023b). SIGN-AIR-Implemented SynergIes, data sharing contracts and Goals between transport modes and AIR transportation. `https://sesarju.eu/projects/sign-air`.

Simpson, R. W. (1969). Scheduling and routing models for airline systems. Technical report, Massachusetts Institute of Technology, Flight Transportation Laboratory, Cambridge, MA.

SNCF (2023a). Fréquentation en gare. `https://ressources.data.sncf.com/explore/dataset/frequentation-gares/information/?disjunctive.nom_gare&disjunctive.code_postal`.

SNCF (2023b). SNCF Open Data. `https://ressources.data.sncf.com/pages/accueil/`.

Srisaeng, P., Baxter, G., and Wild, G. (2015). Using an artificial neural network approach to forecast Australia's domestic passenger air travel demand. *World Review of Intermodal Transportation Research*, 5(3) :281–313.

SteadieSeifi, M., Dellaert, N. P., Nuijten, W., Van Woensel, T., and Raoufi, R. (2014). Multimodal freight transportation planning : A literature review. *European Journal of Operational Research*, 233(1) :1–15.

Sun, J., Ellerbroek, J., and Hoekstra, J. M. (2018). Aircraft initial mass estimation using Bayesian inference method. *Transportation Research Part C : Emerging Technologies*, 90 :59–73.

Tan, Y., Li, Y., Wang, R., Mi, X., Li, Y., Zheng, H., Ke, Y., and Wang, Y. (2022). Improving synchronization in high-speed railway and air intermodality : Integrated train timetable rescheduling and passenger flow forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 23(3) :2651–2667.

Tauler, A. and Martin, S. (2021). Observatorio del Ferrocarril en España. Technical report, Informe de la Fundación de los Ferrocarriles Españoles.

Teodorović, D. and Krčmar-Nožić, E. (1989). Multicriteria model to determine flight frequencies on an airline network under competitive conditions. *Transportation Science*, 23(1) :14–25.

Theis, G., Adler, T., Clarke, J.-P., and Ben-Akiva, M. (2006). Risk aversion to short connections in airline itinerary choice. *Transportation Research Record*, 1951(1) :28–36.

Tirachini, A., Hensher, D. A., and Rose, J. M. (2014). Multimodal pricing and optimal design of urban public transport : The interplay between traffic congestion and bus crowding. *Transportation Research Part B : Methodological*, 61 :33–54.

Toqué, F., Khouadjia, M., Come, E., Trepanier, M., and Oukhellou, L. (2017). Short & long term forecasting of multimodal transport passenger flows with machine learning methods. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 560–566, Yokohama, Japan. IEEE.

Tsai, T.-H., Lee, C.-K., and Wei, C.-H. (2009). Neural-network based temporal feature models for short-term railway passenger demand forecasting. *Expert Systems with Applications*, 36(2) :3728–3736.

UIC-International Union of Railways (2023a). MERITS Database. `https://uic.org/passenger/passenger-services-group/merits`.

UIC-International Union of Railways (2023b). World high speed rolling stock. `https://uic.org/IMG/pdf/20180910_highspeed_rolling_stock.pdf`.

van Nes, R., Hamerslag, R., and Immers, L. (1988). *The design of public transport networks*, volume 1202. National Research Council, Transportation Research Board.

Vansteenwegen, P. and Van Oudheusden, D. (2006). Developing railway timetables which guarantee a better service. *European Journal of Operational Research*, 173(1) :337–350.

Veldhuis, J. (1997). The competitive position of airline networks. *Journal of Air Transport Management*, 3(4) :181–188.

Vespermann, J. and Wald, A. (2011). Intermodal integration in air transportation : Status quo, motives and future developments. *Journal of Transport Geography*, 19(6) :1187–1197.

Wickham, R. R. (1995). *Evaluation of forecasting techniques for short-term demand of air transportation*. PhD thesis, Massachusetts Institute of Technology.

Wong, R. C., Yuen, T. W., Fung, K. W., and Leung, J. M. (2008). Optimizing timetable synchronization for rail mass transit. *Transportation Science*, 42(1) :57–69.

Xia, W. and Zhang, A. (2017). Air and high-speed rail transport integration on profits and welfare : Effects of air-rail connecting time. *Journal of Air Transport Management*, 65 :181–190.

Yan, S. and Chen, C.-H. (2007). Coordinated scheduling models for allied airlines. *Transportation Research Part C : Emerging Technologies*, 15(4) :246–264.

Yang, H., Burghouwt, G., Wang, J., Boonekamp, T., and Dijst, M. (2018). The implications of high-speed railways on air passenger flows in China. *Applied Geography*, 97 :1–9.

Yen, J. Y. (1971). Finding the k-shortest loopless paths in a network. *Management Science*, 17(11) :712–716.

Appendices

A Advancing transport data standardisation and harmonisation : Challenges and opportunities, report from SIGN-AIR workshop

SIGN-AIR is a SESAR project that was launched in the beginning of 2023. The topic of the project is the following one : "*The project will develop and pilot a new platform for the sharing of data in multimodal travel. The platform will provide the means for transport service providers (TSPs) to register, reach data sharing agreements with other TSPs and manage their contractual relationships. The project will address contract templates to simplify the legal management, the electronic management and information provision about each specific contract, routing information for travel companions (TCs) with enriched information about the specific contracts for their customers. The ultimate aim is to facilitate single ticketing through a comprehensive understanding of the contracts and the data managed, among others.*" On November 27^{th}, 2023, the consortium of the SIGN-AIR project organised a workshop, with the objectives of collecting feedback from experts on transportation data, in particular, on the standardisation and harmonisation of TSPs data. This section therefore summarises the information that was shared during this workshop. As an introduction, it has been noticed that data sharing is of major importance for air transportation. Operating a safe flight rely on the exchange of massive data between aircraft, airport, air traffic control systems, etc. The Airport Collaborative Decision Making (CDM) effort demonstrates the potential of data sharing to improve the operation by accelerating processes. The CDM concept relies on information exchange between air transportation stakeholders (airport, airlines, air traffic controllers), to improve the efficiency of airport operations, such as more accurate prediction of aircraft turnaround times and aircraft departure time. It is now implemented in 32 airports in Europe including the largest ones such as Amsterdam-Schiphol, Paris-CDG, and Frankfurt. In 2014, Laplace *et al.* (2014) proposed to extend the concept of CDM by including a passenger perspective and ground transportation modes in the decision process.

In that context, a review of transportation data was conducted during the workshop. According to the Directive 2010/40 of the European Commission on Intelligent Transpor-

tation System, data standards must be defined to ensure interoperability. To achieve that, the European Commission developed several data standard that can be used by public transport operators, in order to share their data. The main standards are described in Figure 9. Transmodel provides accurate definitions of public transportation elements (what

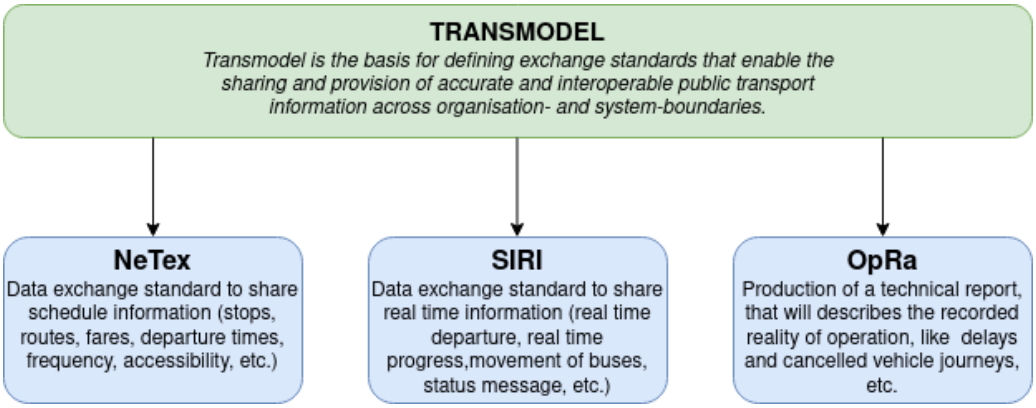


Figure 9 – European data standards.

is a stop point, a vehicle journey, a route, etc.) to help public transportation providers to harmonise their data and facilitate further data exchanges. From these definitions, several data exchange standards were developed such as NeTex, for schedule data, SIRI, for real-time data, and OpRa, for operational reports. A detailed explanation of each format can be found on the Data4PT project website (Data4PT, 2023). In addition, to the data standard, the directive 2010/40 stipulates that each member state of the European Union must establish a National Access Point (NAP) allowing to organise and access transportation data. The list of NAPs can be found in European Commission (2023a).

Despite these initiatives and regulation from the European Commission, TSPs do not share their data easily. Several possible reasons for that were highlighted in the discussion :

— The cost of standardisation : data management has a cost, big data infrastructure must be deployed such as work force to process, and be responsible of the data.

— Heavy standardisation procedure : if the company has the infrastructure and the resources to manage data, it might use data under its own specification and format. It therefore requires to change the way it used to work for years, and some companies may not be ready for this change.

— Cost to share the data without any expected benefit. TSP may not seen a personal interest in gathering data, and are scared about missing a market opportunity.

For these reasons, there is still a gap in finding usable data. While regulations exist, their implementation remains challenging. For instance, the NAP in France for air transportation only provides data of the national airline AirFrance and TUI Fly Belgium, despite many other companies operating in France.

Above data standardisation, the next stage of the discussion was to analyse challenges and opportunities in data sharing. Indeed, assuming that data is harmonised and easily accessible, several challenges remain. First validation phases must be conducted to assess data relevance. If future cooperation should arise, one must ensure that data is not outdated and valid. Regulation must also be provided to TSPs in case of several data sources, to know which data should be included in their operations. A participant of the workshop mentioned the Überlingen crash : two aircraft collapse because they followed different instructions, one from the air traffic controller and another from the TCAS, to avoid the collision. This example highlights the need when cooperating to have the same information between partners. Another challenge that was highlighted is the international coordination. Today, NAPs are individual to each country, and it remains a laborious work to obtain data at the European level all at once. Cross-border coordination is also relevant in Europe, country sizes are small and using a combination of air and rail to travel only makes sense when travelling across several countries. Finally, the SIGN-AIR platform proposes to establish smart contract between TSP, such as single ticket for instance. One of the main challenges in such circumstances, is the clearing. Clearing in finance is defined as the correct and timely transfer of funds to the seller, and securities to the buyer. In context of multimodality, financial agreements between partners should be found, either to assure TSP revenue but also to protect passengers in case of disruptive event.

Finally, opportunities for data sharing and harmonisation were discussed. First, the potential to detect sticking point, especially in terms of timetable synchronisation. Indeed, public transport services generally stops during the night, passengers may therefore use their own car or a taxi to reach or leave the airport in the early morning or in the late evening. In addition, a platform such as SIGN-AIR could benefit transportation suppliers to analyse the demand. If multimodal trips are offered to passengers, some legs will become unprofitable for TSP. Moreover, removing these unused legs could allow TSP to create additional service and offer them to passenger, with the cost saved by deleting the previous one.

This workshop highlighted the valuable initiative of the European Commission to develop regulations and standards regarding public transportation data. These standards should facilitate data sharing and collaboration between transportation stakeholders, resulting in improvement of the passenger door-to-door travel. Unfortunately, the cost of standardisation and the uncertain benefit for transportation operators are a barrier of data sharing, especially for the air transportation sector where competition between airlines remains strong. Assuming that data was easily accessible, several challenges in terms of financial agreement and data validation should also be overcome. However, as illustrated with the CDM concept, data sharing has a large potential in enhancing TSP operations, and improving the passenger door-to-door experience.

B Maximum number of movements per airport

In this section, Tables 6 and 7 respectively present the maximum number of flight departures and flight arrivals, that can scheduled per hour at the 18 airports considered in the thesis. Table 8 present the total number of movements (either departures or arrivals), that can be scheduled per hour at the 18 airports considered (represented here with there ICAO identification code).

Table 6 – Maximum number of departures (flights) that can be scheduled per hour, at the 18 airports considered.

Hour	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
LFPG	25	25	20	20	20	25	38	66	62	64	67	63	65	70	63	65	63	62	62	62	64	53	41	30
LFMN	12	12	12	12	12	12	12	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	12
LFLL	12	12	12	12	12	12	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	12
LFML	7	7	7	7	7	7	15	15	22	22	22	22	22	22	22	22	22	15	15	15	15	15	15	15
LFPO	0	0	0	0	0	0	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	6
LFBO	12	12	12	12	12	12	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	12
EDDF	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
EDDM	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58
EDDT	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
EDDL	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36
EDDH	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31
EDDS	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
LEMD	20	20	20	20	20	29	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52	52	22	20
LEBL	24	24	24	24	24	30	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	22	22
LEMG	24	24	24	24	24	24	24	24	25	25	25	25	25	25	25	25	25	25	25	25	24	24	24	24
LEPA	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34	34
GCLP	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
GCTS	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21

Table 7 – Maximum number of arrivals (flights) that can be scheduled per hour, at the 18 airports considered.

Hour	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
LFPG	30	20	20	20	20	30	41	50	62	62	60	61	54	55	58	56	56	61	60	63	56	53	46	40
LFMN	12	12	12	12	12	12	12	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	12
LFLL	12	12	12	12	12	12	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	12
LFML	5	5	5	5	5	5	9	9	22	22	22	22	22	22	22	22	22	22	18	18	18	18	9	9
LFPO	0	0	0	0	0	0	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	3
LFBO	12	12	12	12	12	12	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	12
EDDF	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
EDDM	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58	58
EDDT	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
EDDL	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
EDDH	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31
EDDS	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
LEMD	20	20	20	20	20	19	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	28	20
LEBL	24	24	24	24	24	18	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	26	26
LEMG	24	24	24	24	24	24	24	24	25	25	25	25	25	25	25	25	25	25	25	25	24	24	24	24
LEPA	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33
GCLP	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
GCTS	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21

Table 8 – Maximum number of departures and arrivals (flights) that can be scheduled per hour, at the 18 airports considered.

Hour	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
LFPG	40	40	32	32	32	40	67	103	109	120	111	112	111	110	105	108	107	107	111	109	108	97	80	62
LFMN	22	22	22	22	22	22	22	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	22
LFLL	24	24	24	24	24	24	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	24
LFML	9	9	9	9	9	9	15	15	44	44	44	44	44	44	44	44	44	44	26	26	26	26	15	15
LFPO	0	0	0	0	0	0	30	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	45	6
LFBO	12	12	12	12	12	12	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	12
EDDF	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106
EDDM	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
EDDT	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78
EDDL	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47
EDDH	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48
EDDS	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48	48
LEMD	38	38	38	38	38	48	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	38
LEBL	48	48	48	48	48	48	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	78	48	48
LEMG	37	37	37	37	37	37	37	37	46	46	46	46	46	46	46	46	46	46	46	46	46	37	37	37
LEPA	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66	66
GCLP	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36
GCTS	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33

C Publications

This appendix lists the scientific publications made along the thesis.

Peer-reviewed conference papers :

— Buire, C., Scozzaro, G., Marzuoli, A., Feron, E., and Delahaye, D. (2021). A year into the pandemic : A passenger perspective on its impact at Paris-Charles de Gaulle airport. In *2021 IEEE International Conference on Big Data*, pages 2925–2935. IEEE

— Buire, C., Delahaye, D., Marzuoli, A., Feron, E., and Mongeau, M. (2022). Air-rail timetable synchronization for a seamless passenger journey. In *Proceedings of International Workshop on ATM/CNS*, pages 79–86, Tokyo, Japan

— Buire, C., Delahaye, D., Mongeau, M., Marzuoli, A., Bueno-Gonzalez, J., Artime, R., G. Cantú Ros, O., and Burrieza-Galán, J. (2023). Leveraging passengers' mobile network data for an integrated air-rail frequency planning in Spain. In 13^{th} *SESAR Innovation Days*, Sevilla, Spain

— Scozzaro, G., Buire, C., Delahaye, D., and Marzuoli, A. (2023). Optimizing air-rail travel connections : A data-driven delay management strategy for seamless passenger journeys. In 13^{th} *SESAR Innovation Days*, Sevilla, Spain

Journal paper :

— Buire, C., Marzuoli, A., Delahaye, D., and Mongeau, M. (2024). Air–rail timetable synchronisation : Improving passenger connections in Europe within and across transportation modes. *Journal of Air Transport Management*, 115 :102526

Data set :

— Buire, C. (2023). Air-Rail Passenger Demand Instances. `https://doi.org/10.57745/5WB9KG`

D Data sources used

Mode	Features	Year	Airport		
			CDG	**FRA**	**MAD**
Air	**Schedule**	2019 and 2021	OAG, Eurocontrol R&D data	OAG, Eurocontrol R&D data	OAG, Eurocontrol R&D data
	Passenger volume		Eurostat	Eurostat	Eurostat
Rail	**Schedule**	2019 and 2021	SNCF open data	Deutsche Bahn, Lockheed (2021)	Renfe
	Passenger volume		×	×	×
Multimodal	**Air-air transfer share**	2019	Maertens *et al.* (2020) (value of 2018)	Fraport 2019	Maertens *et al.* (2020) (value of 2018)
		2021	Aéroport de Paris financial report	Fraport 2021	Burrieza-Galán *et al.* (2022) (value of 2020)
	Air-rail transfer share	2019	SNCF open data	Fraport 2019	Burrieza-Galán *et al.* (2022) (value of 2018)
		2021	SNCF open data	Fraport 2021	Burrieza-Galán *et al.* (2022) (value of 2020)

Table 9 – Raw data sources (green) and data used to estimate (blue) input parameters of the models presented in the thesis.