Leveraging passengers' mobile network data for an integrated air-rail frequency planning in Spain

Clara Buire, Daniel Delahaye, Marcel Mongeau
ENAC, Université de Toulouse
Toulouse, France
clara.buire@enac.fr

Aude Marzuoli
Georgia Institute of Technology
Atlanta, United States
aude.marzuoli@gmail.com

Jerónimo Bueno-Gonzalez, Rubén Artime, Oliva G. Cantú Ros, Javier Burrieza-Galán
NOMMON
Madrid, Spain
jeronimo.bueno@nommon.es

***Abstract*—This study addresses the integrated air-rail frequency planning problem with the goal of estimating jointly flight and long-distance train frequencies while considering potential synchronisation between the two modes and passenger preferences. It introduces a generalised cost model to capture passenger travel preferences and incorporates CO_2 emissions modelling to minimise the environmental impact. The model is tested on the Spanish air-rail transportation network using real passenger demand from mobile phone data. The results indicate that integrating air and rail frequency planning, along with considering environmental costs, can reduce CO_2 emissions by over 66%, thanks to the air-to-rail transfer, with only a 20-minute increase in average door-to-door travel time.**

***Keywords*—Air-rail transportation network, Service network design, Passengers, Mobile network data**

I. Introduction

The European Commission sets two main objectives for the transportation system in 2050: providing passengers with an efficient and resilient multimodal transportation system, and significantly reducing the transportation CO_2 emission. With the goal that 90% of travellers can travel door-to-door in Europe within four hours, the air transportation system should no longer be isolated, and airports should turn into multimodal hubs [1]. At the same time, the Sustainable and Smart Mobility Strategy report [2] claims to reduce CO_2 emissions of the transportation system. In particular, trips within 500 km should be carbon neutral in Europe, and high-speed rail service must be developed for short-haul distances. Coordination mechanisms between air and ground transportation systems are therefore essential to ensure a high level of service for passengers.

In this paper, we propose a modelling framework to design an integrated air and rail transportation network on a domestic market. More precisely, we aim to estimate optimal flight and train frequencies on all the routes of the considered transportation network, including multimodal solutions for passengers. In the following, we refer to this problem as the Air-Rail Service Network Design (ARSND) problem. The model proposed consists in determining train and flight frequencies, minimising the average passenger door-to-door travel time, considering station processing times and transfer times. In addition, passenger preferences are taken into account through a generalised cost that includes sensitivity to travel time, price, and CO_2 emission. We also consider the global environmental cost of the transportation network. The model is tested on the Spanish transportation network, using passengers' mobile network data to build the Origin-Destination (OD) demand matrix.

The paper is organised as follows. Section II describes the ARSND problem, provides a brief literature review on air-rail coordination, and details the transportation network and passenger demand models. Section III presents the mathematical formulation of the problem, and the resolution approach. Finally, Section IV presents the Spanish case study, and discusses the results obtained.

II. Integrated air-rail service network design

A. Previous studies

Air and rail have long been seen as competitors, especially on the short-haul market [3]–[5]. However, with increasing airport congestion and environmental awareness, recent studies show that high-speed trains are a good alternative to short-haul flights, especially in view of relieving airport congestion [6], [7]. To this end, the European Commission launched several research projects regarding multimodality. IMHOTEP [8] proposes to extend airport operations management, including collaboration with ground transport modes. In the TRANSIT project [9], coordination mechanisms between air and ground transportation modes are implemented to mitigate the impact of a delay on a door-to-door journey and to offer passengers seamless transfers between rail and air. MODUS [10] aims to measure the performance of the transportation system considering the entire passenger door-to-door journey. Indeed, the shift from flight-oriented to passenger-oriented metrics reveals that in-vehicle time may differ significantly from door-to-door travel time [11], [12].

In parallel, some studies examine the potential benefits of integrating air-rail frequency planning problems. The frequency

SESAR Innovation Days
27-30 November 2023
Seville

planning problem, which relies in determining the daily volume of flights or trains, is traditionally made individually by each supplier, considering potential competition, several months before the day of operations [13]–[16]. Okumura and Tsukai [17] propose an air-rail integration by determining the optimal flight frequencies and train speeds under various airport capacity and rail-line length constraints. However, the author assumes that train frequencies remain the same as initially scheduled. Allard and Moura [18] study air-rail passenger network design. These authors propose a model to design a hub-and-spoke transportation network including both air and rail.

Here, we propose to address simultaneously the classical frequency planning problems of airlines and railway operators. Both the flight and rail frequencies are estimated. In addition, we do not impose a hub-and-spoke network topology and allow point-to-point services when relevant. Passengers' preferences are extracted from the analysis of mobile phone data, and the environmental cost of the transportation network is also considered in the objective function.

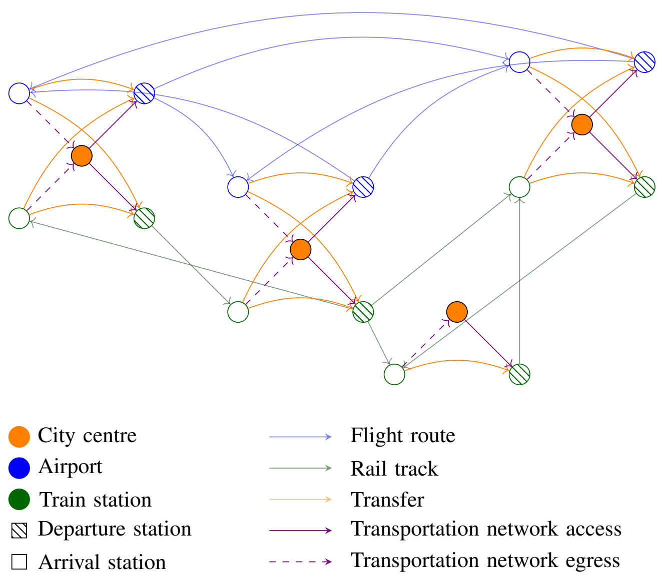


Figure 1. Illustration of air-rail multimodal network: 4-city example.

B. *Multimodal transportation network model*

The transportation network is represented by an oriented graph, $G = (V, E)$, where the vertex set V is the set of nodes and the edge set E 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 E is partitioned into three subsets: E^{air} represents flight routes between airports, E^{rail} corresponds to rail tracks, and E^{transfer} models transfers between stations or city centres. In addition, each airport and a subset of train station node is duplicated into one arrival node and one departure node. This way one can model a transfer between two consecutive flights (respectively, trains) at the same airport (respectively, train station). Transfers can also occur between two different modes, for instance between an airport and a train station if a passenger catches a train after a flight. Finally, the edge set E also includes edges from city centre nodes to transportation stations (*accessing edges*) and edges from stations to city centres (*egressing edges*). An illustration of the considered network is presented in Figure 1.

For each edge $e \in E$, several costs are defined: the travel time, the price, and the CO_2 emission.

1) Travel time cost: For flight-route and rail-track edges, the travel time is estimated as the average travel time by flight and train initially scheduled by transportation suppliers. Regarding transfer edges, an average transfer time, noted t^{transfer}, is considered. If passengers must shift between two stations, the travel time between the two stations is included to the average transfer time. This travel time is computed by multiplying the great circle distance between the two stations with an average speed of 90km/h. If the stations are located at the same place (*e.g.*, if the train station is directly 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 edges and egress edges. 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 [19] that airport processing time and passengers conservative behaviour may significantly increase the door-to-door journey. We therefore add an average outbound processing time (t^{dep}), to the access edge travel time, and an average inbound processing time (t^{arr}) to the egress edge travel time. The station processing times are therefore included in the total door-to-door travel time.

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, 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 that issue, we proceed as follows:

- We only consider one 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 [20]. 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 function, the fuel consumption per kilometre can be estimated: the total amount of fuel (in kg) for a scheduled flight is 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 [21].

The CO_2 emission of trains per km is provided by the Office of Rail and Road [22]. The CO_2 emission of a train is estimated as the product of the emission per kilometre and the distance travelled by train. For transfers (access and egress edges) the CO_2 emission is assumed to be null.

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 can be obtained from [23] and [24], respectively.

C. Passenger demand model

The passenger demand corresponds to the number of passengers that want to travel between each OD pair. The OD demand is represented by a set of *commodities* C. A commodity $c \in C$ is a tuple (O_c, D_c, d_c), where O_c is the origin, D_c is the destination and d_c is the number of passengers. In the following, we name a *path* a sequence of edges. For instance, for two connected cities with airports, a path corresponds to: the access edge from the city centre to the airport, the flight edge, and finally the egress edge from the airport to the destination city. For each commodity $c \in C$, we denote $\mathcal{P}_c$ the set of paths that connect O_c to D_c.

Passengers may have various travel preferences: the business travellers are likely to be more concerned about the travel duration than leisure travellers are [25]. 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 for the cost 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 follows:

$$\tau_{cp} = \sum_{i \in I} \alpha_{ci} \tau_{pi}, \tag{1}$$

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:

$$\tau_{pi} = \sum_{e \in p} c_{ei}, \tag{2}$$

where c_{ei} is the cost i of edge $e \in E$, as described above.

III. Mathematical formulation and resolution approach

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

A. 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 we denote $\mathcal{P}_{c,e}$ the set of paths using edge e, $e \in E$. A *train-line* is a sequence of stations served by the same train. Let L be the set of train-lines. We assume that a train should provide service to the entire line. We further define for each edge $e \in E$, the set of lines using e, noted L_e. For simplicity, this preliminary study assumes that the air and rail fleets are composed of exactly one aircraft type and one train type. For any edge $e \in E^{\text{air}} \cup E^{\text{rail}}$, the capacity of a vehicle (aircraft of rail-car) is known and denoted C_e and the cost c_{ei} is given for each cost $i \in I$.

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 can be built. In the following, we denote f^{air} and f^{rail} the maximum flight and train frequencies, per flight route and rail track, to take into account capacity constraints. In addition, we denote F and R the maximum number of flights and trains that can be scheduled for the whole day, respectively.

B. 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 edge $e \in E$, an auxiliary continuous decision variables v_e, giving the number of passengers assigned to travel through edge e. We also define an integer decision variable, y_e, that counts the number of services (flight or train) to schedule on edge $e \in E$. Finally, for each train line $l \in L$, we define an integer optimisation variables f_l that decides the frequency of l.

C. Objective function and constraints

The aim is to route passengers in the network according to their preferences while minimising the CO_2 emissions of the transportation network. The problem is formulated as a bi-criterion optimisation problem where the functions to be minimised are the following:

$$F_1(x, v, y, f) = \sum_{c \in C} d_c \sum_{p \in \mathcal{P}_c} x_{cp} \tau_{cp}, \tag{3}$$

which represents the passenger preferences, and:

$$F_2(x, v, y, f) = \sum_{e \in E} y_e c_{e,\text{CO}_2}, \tag{4}$$

which is the global CO_2 emission criterion. We propose the following Mixed-Integer Linear Programming (MILP) formulation:

$$\min_{x,v,y,f} (F_1, F_2), \tag{5}$$

$$
\begin{aligned}
&\sum_{p\in\mathcal{P}_c} x_{cp} = 1 && c \in C, && \text{(5a)}\\
&\sum_{c\in C_e}\sum_{p\in\mathcal{P}_{c,e}} x_{cp} d_c = v_e && e \in E, && \text{(5b)}\\
&v_e \le C_e y_e && e \in E, && \text{(5c)}\\
&y_e \le f^{\text{air}} && e \in E^{\text{air}}, && \text{(5d)}\\
&y_e \le f^{\text{rail}} && e \in E^{\text{rail}}, && \text{(5e)}\\
&y_e = \sum_{l\in L_e} f_l && e \in E^{\text{rail}}, && \text{(5f)}\\
&\sum_{e\in E^{\text{air}}} y_e \le F, && && \text{(5g)}\\
&\sum_{l\in L} f_l \le R, && && \text{(5h)}\\
&1 \le f_l && l \in L, && \text{(5i)}\\
&x_{cp} \in [0,1] && c \in C, p \in \mathcal{P}_c, && \text{(5j)}\\
&v_e \in \mathbb{R} && e \in E, && \text{(5k)}\\
&y_e \in \{0,1,\ldots,f^{\text{air}}\} && e \in E^{\text{air}}, && \text{(5l)}\\
&y_e \in \{0,1,\ldots,f^{\text{rail}}\} && e \in E^{\text{rail}}, && \text{(5m)}\\
&f_l \in \{0,1,\ldots,f^{\text{rail}}\} && l \in L. && \text{(5n)}
\end{aligned}
$$

Constraints (5a) ensure that all passengers are routed in the network. Constraints (5b) count the number of passengers routed through edge e. Constraints (5c) implement the capacity bound on each edge while constraints (5d) and (5e) implement the upper bounds on frequencies for each edge. Constraint (5f) ensures that a train is scheduled for an entire line. Constraints (5g) and (5h) implement upper bounds on the total number of scheduled flights and trains. Constraint (5i) guarantees a minimum level of service for every *train-line*. Finally, constraints (5j)-(5n) specify the definition domain of the decision variables.

D. Resolution approach

To obtain a realistic solution and reduce the computation time, a pre-processing step is made. For each commodity c, there are several possible paths to travel from O_c to D_c. However, one can assume that passengers will prefer to travel through shortest paths, according to the cost defined above. Hence, 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. This computation is made using Yen algorithm [26], and the value of the parameter k is set by the user.

The bi-criterion optimisation problem is addressed via a weighted-sum scalarisation of the two criteria: the problem is rewritten as a mono-criterion optimisation problem:

$$\min_{x,v,y,f} \lambda F_1 + (1-\lambda) F_2 \qquad (6)$$

where λ is a user-defined parameter. The resolution of the optimisation problem formulation is made using the MILP solver Gurobi [27], version 9.1.2. The global resolution framework is presented in Figure 2. Note that if λ parameter

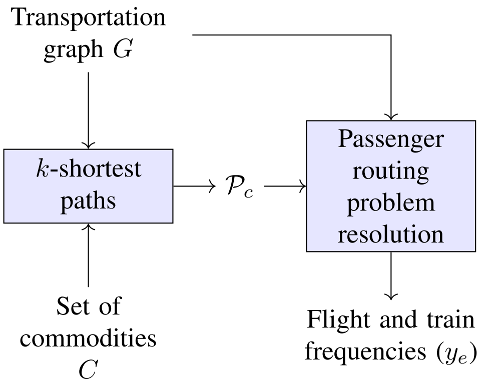


Figure 2. Resolution framework of the ARSND problem.

is set to 1, the CO_2 global criterion is not considered in the objective function. In this case, there may be multiple solutions minimising the total passenger travel time, 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 simply do the following post-processing set: we launch the solver one last time by constraining the value of F_1 to be as good as previously found, while minimising the total frequency.

IV. 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.

A. Spanish transportation network

Train and flight schedules data are collected for January 20, 2023 from the website of RENFE [28], which is the most important rail operator in Spain, and from OAG [29], 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. This aircraft type is then used to model CO_2 emission according to the aircraft characteristics [30]. Regarding rail CO_2 emission, rail services are served by electric trains in Spain [31]. The value of 358g CO_2/km for an electric train, computed by the Office of Rail and Road [22], 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 is set to 500 passengers, and the flight average capacity to 189, as this is the value found from

TABLE I. Average transfer and processing times at stations (minutes).

	t^{dep}	t^{arr}	t^{transfer}
Airport	90	20	60
Train station	10	0	15

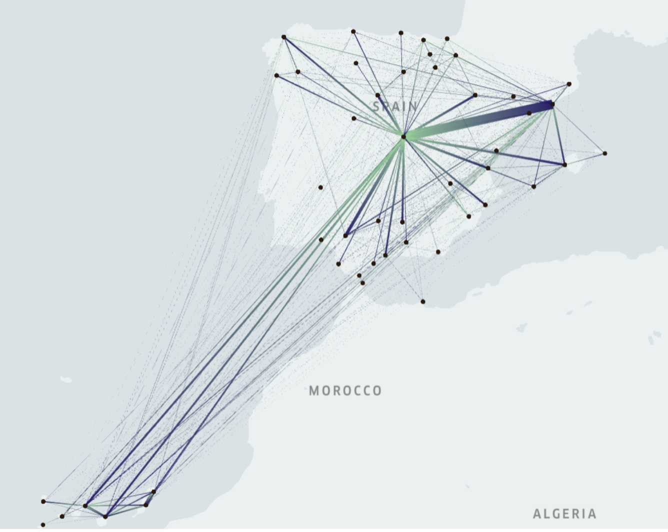


Figure 3. OD matrix demand and the 48 cities considered. Bolder edges correspond to higher number of passengers.

OAG for the Boeing 737-800 on that day. A filling rate of 100% is assumed on each mode.

For simplicity, we restrict the node set to 48 cities of Spain, including the largest ones and at least one city per island. These cities are represented on Figure 3. The values of the parameters set for the study are presented in Table I.

B. *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. MND consist of all the interactions between mobile devices and the antennas of the MNO. The data are analysed using a processing pipeline for reconstructing door-to-door passenger journeys [32]. 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 paper, 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 district level between the 48 cities selected, covering only trips above 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 3 displays the travel demand between the 48 cities. In total, more than 97,000 people travel on that day, with a majority of trips between Madrid and Barcelona. Less than 1% of trips combine air and rail modes, and most of air-rail transfers take place in Madrid, as presented in Table II. In order to consider these transfers, associated train stations are duplicated in the transportation network graph model.

TABLE II. 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

This demand serves as an input for the optimisation problem.

C. *Optimisation results*

Computations are performed on a laptop equipped with an AMD Ryzen 5 4500U CPU and 16 GB RAM. The computation time limit is set to two hours and the MIP gap to 0.5%. The values used for the model are defined according to the initial schedules and are summarised in Table III. In a first step, only travel time is considered in the passengers' generalised cost function ($I = \{\text{Travel time}\}$).

TABLE III. PARAMETERS OF THE CASE STUDY.

Parameter	f^{air}	f^{rail}	F	R	k
Value	25	50	1033	1584	20

1) Pool of solutions: Computation information for several values of λ are summarised in Table IV. Figure 4 displays the value 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

TABLE IV. COMPUTATION INFORMATION.

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
04	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

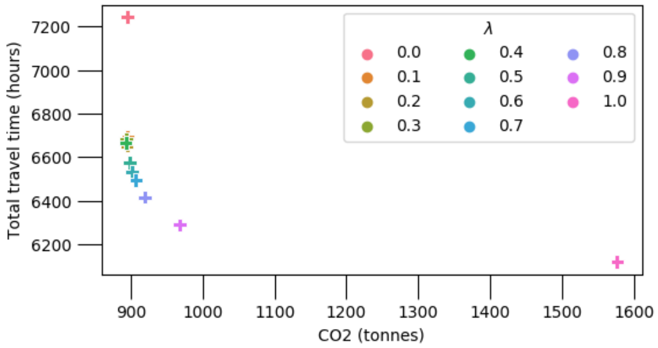


Figure 4. Objective function criteria as a function of λ parameter.

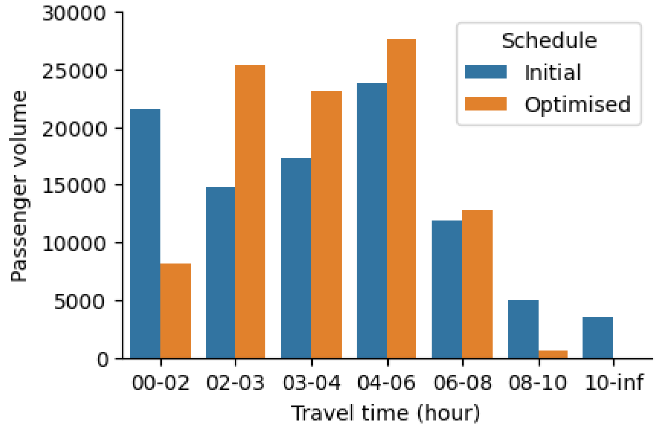


Figure 5. Passenger volume per travel time window.

the objective function increases (Equation 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 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 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 be reduced but the CO_2 cost for travelling by road is not equal to zero. In the following, the λ parameter is set to 0.9.

2) Passenger trips: In the remaining of the study, results are compared with the initial supply and demand data.

Table V summarises the number of passengers per mode. In the initial planning, only 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 by using 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 one these routes is not relevant. In total, 57% of passengers are not affected by the synchronisation.

Figure 5 presents the distribution of the total travel times for both the initial and optimised plannings. Note that for the initial schedule, travel times are obtained with mobile phone data. For January 26, 2022, 24% of trips above 50 km using public transportation last between four and six hours. This duration remains the most represented one 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. 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 V.

TABLE V. 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

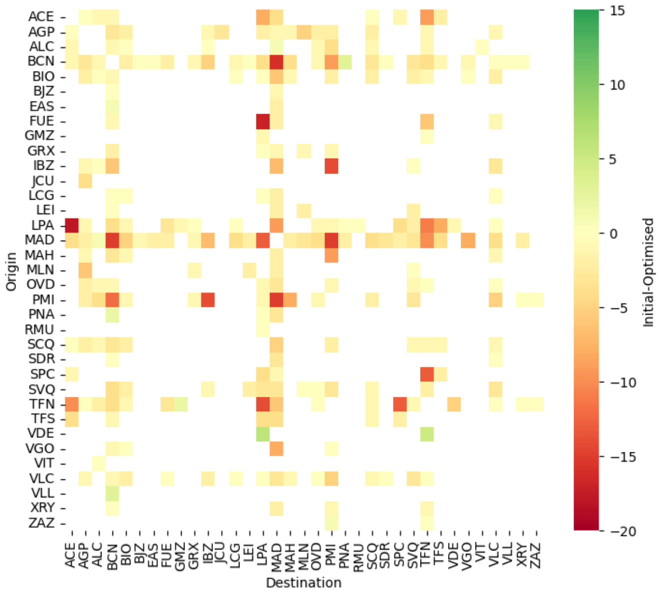
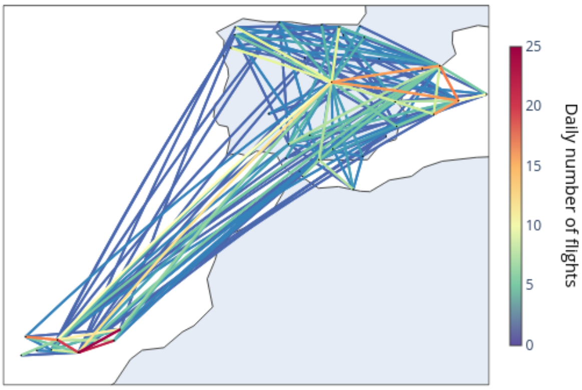
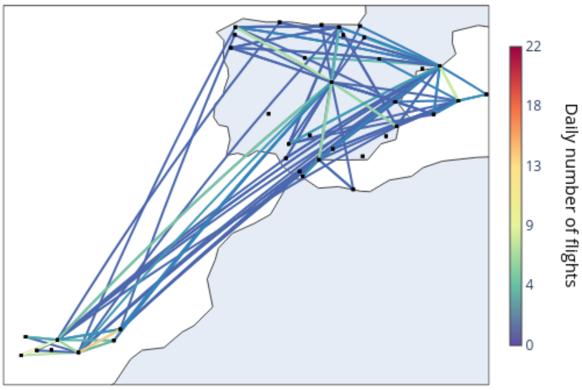


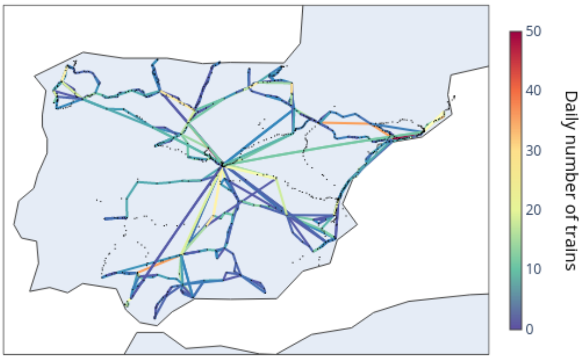
Figure 6. Reduction in the number of daily flights per OD pair.



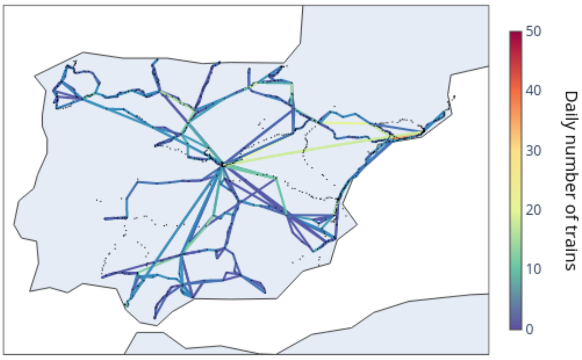
(a) Flight frequencies - Initial schedule



(b) Flight frequencies - Integrated schedule



(c) Train frequencies - Initial schedule



(d) Train frequencies - Integrated schedule

Figure 7. Flights (top line) and train (bottom line) frequencies of the initial (left column) and optimised (right column) plannings.

3) Integrated transportation network: Figure 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. Results show a significant reduction in the number of daily flights between Barcelona and Madrid. Actually, there are 30 fewer flights on this OD segment (considering both directions) after optimisation. This decrease can be attributed to the cost-efficiency of train travel, both in terms of carbon emissions and total door-to-door travel time. Comparatively, taking the train requires 170 minutes, whereas flying takes 205 minutes. Moreover, for air connecting passengers, the environmental impact of these flights outweighs the marginal time savings in reaching Madrid airport from Barcelona via train. The solution thereby routes all passengers by rail on that segment. Note, however, that we did not consider the discomfort induced from shifting from the train station to the airport in Madrid, since the train station is not directly at the airport. Transportation operators can therefore ensure a minimum flight frequency between Barcelona and Madrid to account for connecting passengers.

TABLE VI. TOTAL FLIGHT AND TRAIN FREQUENCIES.

	Initial schedule	Optimised schedule
Flights	1033	498
Trains	1584	619

One can also observe a large reduction in the number of flights between Spanish islands airports (LPA, ACE, IBZ, PMI, TNF, FUE). Figure 3 reveals that a small number of passengers travelled between Spanish islands on that day, maybe due to the winter season. In addition, direct flights from mainland to islands are now scheduled only from a small number of cities. In particular, Figure 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.

The total train and flight frequencies for the initial and optimised planning are summarised in Table VI. For $\lambda = 0.9$, 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 considered 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 taken into account, 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 train, with a limited impact on passengers door-to-door journey, as demonstrated on the Madrid-Barcelona segment.

V. Conclusion and future works

This paper introduces a Mixed-Integer Linear Problem formulation of the air-rail integrated frequency planning problem that considers both passenger perspective and CO_2 emissions. The frequency planning model proposed handle passenger preferences such as travel time, price or environmental awareness. Furthermore, travel time is estimated door-to-door, and not from one station to another, including 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 an input for the optimisation model. Results show that considering CO_2 emission while designing long-distance schedule succeeds in reducing by several tonnes the carbon footprint of the transportation system, at the expense of an increase of the average door-to-door travel time of passengers by only 20 minutes. 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.

Future works 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 are 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, CO_2 emission model is developed as a first estimator but it can be improved by including fleet scheduling in the optimisation process. Finally, one of the future research tracks is to extend this model by considering time of the day. Indeed, the frequency can be estimated per time-window intervals considering an expected passenger arrival time at destination.

Acknowledgement

This project has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114815 under European Union's Horizon Europe research and innovation programme, and from the SESAR Joint Undertaking under grant agreement No 893209 under European Union's Horizon 2020 research and innovation programme.

References

[1] European Commission and Directorate-General for Mobility and Transport and Directorate-General for Research and Innovation, "Flightpath 2050 : Europe's vision for aviation," Tech. Rep., 2011.

[2] European Commission, Directorate-General for Mobility and Transport, "Sustainable and smart mobility strategy – Putting European transport on track for the future," European Commission, Tech. Rep., 2020.

[3] Y. Park and H.-K. Ha, "Analysis of the impact of high-speed railroad service on air transport demand," *Transportation Research Part E: Logistics and Transportation Review*, vol. 42, no. 2, pp. 95–104, 2006.

[4] C. Behrens and E. Pels, "Intermodal competition in the London–Paris passenger market: High-speed rail and air transport," *Journal of Urban Economics*, vol. 71, no. 3, pp. 278–288, 2012.

[5] J. L. Jiménez and O. Betancor, "When trains go faster than planes: The strategic reaction of airlines in Spain," *Transport Policy*, vol. 23, pp. 34–41, 2012.

[6] M. Givoni and D. Banister, "Airline and railway integration," *Transport Policy*, vol. 13, no. 5, pp. 386–397, 2006.

[7] C. Jiang and A. Zhang, "Effects of high-speed rail and airline cooperation under hub airport capacity constraint," *Transportation Research Part B: Methodological*, vol. 60, pp. 33–49, 2014.

[8] M. M. Mota, P. Scala, R. Herranz, M. Schultz, and E. Jimenez, "Creating the future airport passenger experience: IMHOTEP," in *European Modelling Simulation Symposium, Athens, Greece*, 2020.

[9] J. Bueno, J. Burrieza, O. García Cantú, C. Livingston, M. Balac, G. Scozzaro, and C. Buire, "TRANSIT final project results report," 2022.

[10] A. Paul, "Modelling and assessing the role of air transport in an integrated, intermodal transport system," https://modus-project.eu/, 2020.

[11] A. Cook, G. Tanner, S. Cristóbal, and M. Zanin, "Passenger-oriented enhanced metrics," *Second SESAR Innovation Days*, 2012.

[12] P. Monmousseau, D. Delahaye, A. Marzuoli, and E. Feron, "Door-to-door travel time analysis from Paris to London and Amsterdam using Uber data," in *SID 2019, 9th SESAR Innovation Days*, 2019.

[13] M. R. Bussieck, P. Kreuzer, and U. T. Zimmermann, "Optimal lines for railway systems," *European Journal of Operational Research*, vol. 96, no. 1, pp. 54–63, 1997.

[14] T. L. Magnanti and R. T. Wong, "Network design and transportation planning: Models and algorithms," *Transportation science*, vol. 18, no. 1, pp. 1–55, 1984.

[15] R. W. Simpson, "Scheduling and routing models for airline systems," Massachusetts Institute of Technology, Flight Transportation Laboratory, Tech. Rep., 1969.

[16] A. Ghobrial, N. Balakrishnan, and A. Kanafani, "A heuristic model for frequency planning and aircraft routing in small size airlines," *Transportation Planning and Technology*, vol. 16, no. 4, pp. 235–249, 1992.

[17] M. Okumura and M. Tsukai, "Air-rail inter-modal network design under hub capacity constraint," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 7, pp. 180–194, 2007.

[18] R. F. Allard and F. M. M. V. e. Moura, "Optimizing high-speed rail and air transport intermodal passenger network design," *Transportation Research Record*, vol. 2448, no. 1, pp. 11–20, 2014.

[19] Innaxis, "Kerb-to-gate travel time distribution," http://visual.innaxis.org/dataset2050/d2d-time-distribution/.

[20] J. Sun, J. Ellerbroek, and J. M. Hoekstra, "Aircraft initial mass estimation using Bayesian inference method," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 59–73, 2018.

[21] International Air Transport Association, "IATA Carbon Offset Program," https://www.iata.org/contentassets/922ebc4cbcd24c4d9fd55933e7070947/icop_faq_general-for-airline-participants.pdf, 2022.

[22] Office of Rail and Road, "Rail emissions," https://dataportal.orr.gov.uk/statistics/infrastructure-and-emissions/rail-emissions/, 2022.

[23] C. Livingston, M. Balac, A. Gregg, J. Bueno, J. Burrieza, G. Scozzaro, and C. Buire, "Impact Assessment of New Intermodal Concepts and Passenger Information Services: Conclusions and Recommendations," TRANSIT Consortium, Tech. Rep., 2022.

[24] A. Tauler and S. Martin, "Observatorio del Ferrocarril en España," Informe de la Fundación de los Ferrocarriles Españoles, Tech. Rep., 2021.

[25] E. Pels, P. Nijkamp, and P. Rietveld, "Access to and competition between airports: a case study for the San Francisco Bay area," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 1, pp. 71–83, 2003.

[26] J. Y. Yen, "Finding the k shortest loopless paths in a network," *Management Science*, vol. 17, no. 11, pp. 712–716, 1971.

[27] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," https://www.gurobi.com, 2023.

[28] RENFE, "Renfe Data," https://data.renfe.com/dataset?res_format=GTFS, 2023.

[29] OAG, "Schedule analyser," https://www.oag.com/schedules-analyser.

[30] Boeing, "Next-generation 737 airplane characteristics for airport planning," https://www.boeing.com/resources/boeingdotcom/commercial/airports/acaps/737NG_REVA.pdf, 2023.

[31] RENFE, "Train fleet," https://www.renfe.com/es/en/renfe-group/renfe-group/fleet-of-trains, 2023.

[32] J. Burrieza-Galán, R. Jordá, A. Gregg, P. Ruiz, R. Rodríguez, M. Sala, J. Torres, P. García-Albertos, O. Cantú Ros, and R. Herranz, "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*, vol. 100, p. 102163, 2022.