Assignment 1

AE4423 – Airline Planning Optimisation

Group 3

Assignment 1

AE4423 – Airline Planning Optimisation

by

Quirine Japikse (5040930),
Naomi Cornelissen (5105765),
Kristian Terlien (5169941)

TU Delft

Contents

1

1A - Demand forecast

1.1. Formulas used

1. The gravity model for demand between airports i and j:

$$D_{ij} = k\frac{(pop_i\, pop_j)^{b_1}(GDP_i\, GDP_j)^{b_2}}{(f \times d_{ij})^{b_3}}$$

2. The logarithmic formulation:

$$\log(D_{ij}) = \log(k) + b_1 \log(pop_i\, pop_j) + b_2 \log(GDP_i\, GDP_j) - b_3 \log(f\, d_{ij})$$

3. The formula for yearly growth:

$$\text{Yearly Growth} = \frac{2023 - 2020}{3}$$

4. The projection for the year 2025:

$$2025 = 2023 + 2 \times \text{Yearly Growth}$$

5. The formula for the great-circle distance between two airports:

$$d_{ij} = R_E \cdot \Delta\sigma_{ij},$$

where:

$$\Delta\sigma_{ij} = 2 \arcsin \sqrt{\sin^2\left(\frac{\varphi_1 - \varphi_2}{2}\right) + \cos\varphi_1 \cos\varphi_2 \sin^2\left(\frac{\lambda_1 - \lambda_2}{2}\right)}.$$

1.2. Explanation

1.2.1. Distance between airports

To calculate the great-circle distance between two airports, a function based on the haversine formula is implemented. The distance between two points on the Earth's surface is given by the great-circle distance formula (Formula 5). The angular distance $\Delta\sigma_{ij}$ is calculated as described in formula 5.

To calculate the distances for a list of airport pairs, the function retrieves the latitude and longitude coordinates of each airport from the dataset, converts these coordinates from degrees to radians, and applies the haversine formula to compute $\Delta\sigma_{ij}$. The result is then multiplied by R_E to obtain the great-circle distance d_{ij}.

1.2.2. Forecast demand for 2025

To forecast the demand for 2025, the annual growth rates for population and GDP were calculated based on data from 2020 and 2023. These growth rates were used to estimate the population and GDP values for 2025 by assuming a consistent yearly increase. The calculated values for 2025 were then added to the existing dataset for each airport.

Next, the product of the population and GDP values for each airport pair was computed. To ensure consistency with the model requirements, the logarithms of these computed products, along with the logarithm of the distance between airport pairs, were calculated and compiled into a dataset.

Using this dataset, predictions for the demand in 2025 were generated by applying the trained regression model. The predicted values were exponentiated to revert them from the logarithmic scale to the actual demand values. This approach enabled us to leverage historical data trends and extrapolate them to estimate the future demand for 2025.

1.3. Results

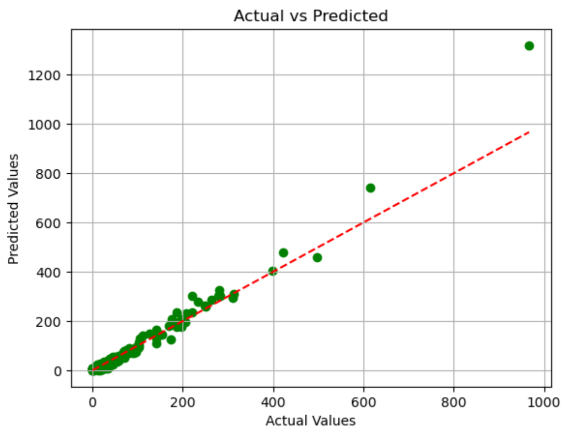

Figure 1.1: Actual vs predicted forecasts for 2025.

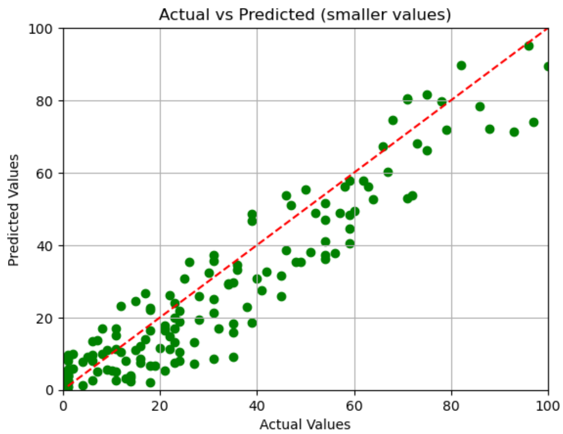

Figure 1.2: Zoomed-in forecasts for 2025.

The first graph (Figure 1.1) illustrates the comparison between the actual and predicted demand values, with the red dashed line representing the ideal case where predicted values perfectly match the actual values. The green points cluster around the red line, indicating a strong correlation between the predictions and the actual demand data. However, a few points, particularly at higher demand values, deviate significantly from the line, suggesting some over- or underestimation in the predictions for extreme cases.

The second graph (Figure 1.2) provides a zoomed-in view of the smaller demand values, revealing a consistent alignment of the green points with the red dashed line. This indicates that the model performs well in forecasting demand within this range. Overall, the results demonstrate that the forecasting model is robust for most scenarios.

2

1B - Network and fleet development

2.1. Mathematical model

The notation used for this mathematical formulation is provided in Table 2.1.

Table 2.1: Notation

Sets and indices	
I	Set of airports $i, j \in I$
K	Set of aircraft types $k \in K$
Parameters	
s_k	seats available in aircraft k
V_k	Speed of aircraft k
TAT_k	Turnaround time for aircraft k in minutes
r_k	Maximum range of aircraft k in km
rr_k	Required runway length for aircraft k
rl_i	Runway length at airport i
c_k	Weekly leasing costs of aircraft k
cx_k	Fixed costs for aircraft k
ct_k	Time-based costs for aircraft k
cf_k	Fuel costs for aircraft k
BT	Total available operation hours per week
LF	Load factor in an aircraft
d_{ij}	distance from airport i to airport j
q_{ij}	demand for flights between airport i and airport j
g_i	Hub indicator (1 for non-hub aiports, 0 for hub airports)
y_{ij}	Yield for origin-destination pair i, j
$CASK_{ijk}$	Cost per available seat kilometer for flights between i and j using aircraft k
Decision variables	
x_{ij}	Direct flow between airport i and j
w_{ij}	Flow transferring through the hub
z_{ijk}	Number of flights between airport i and airport j using aircraft k
ac_k	Number of aircrafts of type k used

Where:

$$y_{EUR_{ij}} = 5.9 d_{ij}^{-0.76} + 0.043 \tag{2.1}$$

$$C^k_{T_{ij}} = c^k_T \frac{d_{ij}}{V^k} \tag{2.2}$$

$$C^k_{F_{ij}} = \frac{c^k_F \times f}{1.5} d_{ij} \tag{2.3}$$

$$C^k_{ij} = C^k_X + C^k_{T_{ij}} + C^k_{F_{ij}} \tag{2.4}$$

With assumptions being:

$$LF = 75\% \tag{2.5}$$

$$C^k_{4j} = 0.7 \times C^k_{ij} \tag{2.6}$$

$$C^k_{i4} = 0.7 \times C^k_{ij} \tag{2.7}$$

The mathematical formulation then follows as:

$$\max \quad \sum_{i\in I} \sum_{j\in I, i\neq j} y_{ij} d_{ij}(x_{ij} + w_{ij}) - \sum_{i\in I} \sum_{j\in I, i\neq j} \sum_{k\in K} C_{ijk} z_{ijk} - \sum_{k\in K} c_k ac_k \tag{2.8}$$

Subject to:

$$x_{ij} + w_{ij} \leq q_{ij}, \qquad \forall i,j \in I, i\neq j \tag{2.9}$$

$$w_{ij} \leq q_{ij} g_i g_j, \qquad \forall i,j \in I, i\neq j \tag{2.10}$$

$$x_{ij} + \sum_{m\in I} w_{im}(1-g_j) + \sum_{m\in I} w_{mj}(1-g_i) \leq \sum_{k\in K} z_{ijk} s_k LF, \qquad \forall i,j \in I, i\neq j \tag{2.11}$$

$$\sum_{j\in I} z_{ijk} - \sum_{j\in I} z_{jik} = 0, \qquad \forall i \in I, k\in K \tag{2.12}$$

$$\sum_{i\in I} \sum_{j\in I} \left(\frac{d_{ij}}{V_k} + TAT_k \big(1.5(1-g_j) + g_j\big) \right) z_{ijk} \leq BT \cdot ac_k, \qquad \forall k \in K \tag{2.13}$$

$$z_{ijk} \leq a_{ijk}, \qquad \forall i,j \in I, k\in K \tag{2.14}$$

$$a_{ijk} = \begin{cases} 10000, & \text{if } d_{ij} \leq r_k \text{ and } rr_k \leq rl_i andrr_k \leq rkl_j \\ 0, & \text{otherwise} \end{cases}, \qquad \forall i,j \in I, k\in K \tag{2.15}$$

$$x_{ij}, w_{ij} \geq 0, \qquad \forall i,j \in I \tag{2.16}$$

$$z_{ijk} \in \mathbb{Z}_{\geq 0}, \qquad \forall i,j \in I, k\in K \tag{2.17}$$

$$ac_k \in \mathbb{Z}_{\geq 0}, \qquad \forall k \in K \tag{2.18}$$

The yield is calculated using formula 2.1. The total operating costs (2.4) is based on the fixed operating costs, the time based costs and fuel costs. The time based costs is calculated with formula 2.2 and the fuel costs are calculated with formula 2.3. The assumptions made in this problem are that the load factor of an airplane is 75% (2.5). Furthermore, flights to and from the hub, Madrid, being indexed 4, is 30% cheaper in the total operating costs, this can be seen in the functions 2.7 and 2.6.

The objective function 2.8 aims to maximize the profit of the airline network. It is composed of three terms: revenue from passenger flows (direct and transferring), costs associated with operating flights, and the weekly leasing costs of aircraft. This objective is subject to the following constraints. The demand constraint 2.9 ensure that the total passenger flow (direct and transferring) between any two airports does not exceed the demand for that route. Additionally, transferring flows are restricted to routes involving the hub airport, which is specified by a hub indicator parameter 2.10. The aircraft capacity constraint 2.11ensures that the total passenger flow on any route does not exceed the available seating capacity of the aircraft assigned to that route, adjusted by the load factor. Flow balance constraints enforce that the number of flights departing from any airport equals the number of flights arriving, seen in constraint 2.12.

Aircraft utilization constraints 2.13 ensure that the total operating hours of each aircraft type, including flight times and turnaround times, do not exceed the available operational time per week. Furthermore, range and runway constraints enforce that flights are only operated if the aircraft can cover the required distance and the runway lengths at the origin and destination airports are sufficient for the aircraft type, seen in constraint 2.14. Finally, non-negativity and integer constraints ensure that passenger flows and the number of flights are realistic, and the number of aircraft used is an integer value 2.16, 2.17 and 2.18.

2.2. Results

The optimal solution yields an objective revenue of € 26314.9. The corresponding flight plan is presented in Table 2.2. The results indicate that the highest number of flights are scheduled from London and Madrid. This can be attributed to Madrid serving as the designated hub, which naturally centralizes operations, and London being a major metropolitan area with significant passenger demand. One remarkable flight in the plan is between Stockholm and Helsinki. Since the two cities are disconnected from the network, there has to be a flight with no passengers to exploit this connection. However, these empty flights to move the aircraft are not included in the model. For this reason, the model could be adapted to account for these flights.

Table 2.2: Flight plan of occurence of round-trip connections

From	To	Frequency per week	Aircraft type
London	Paris	6	3
London	Amsterdam	2	2
London	Amsterdam	3	6
London	Madrid	2	1
London	Madrid	3	2
London	Dublin	3	2
Paris	Amsterdam	3	2
Paris	Madrid	2	1
Amsterdam	Madrid	1	2
Frankfurt	Munich	2	2
Frankfurt	Madrid	1	2
Madrid	Barcelona	5	2
Madrid	Munich	1	3
Madrid	Rome	1	3
Madrid	Palermo	1	2
Madrid	Madeira	1	2
Madrid	Dublin	1	2
Madrid	Lisbon	3	2
Madrid	Berlin	1	3
Dublin	Edinburgh	1	1
Stockholm	Helsinki	1	2
Stockholm	Helsinki	1	3

The weekly flight plan results in a network map which is shown in figure 2.1

In table 2.3, the number of aircraft of each type leased in the optimal solution is shown. The solution indicates that one aircraft of types 2 and 3 will be leased, while no aircraft will be leased for the other two types. This result reflects the balance between demand and cost-effectiveness, as the demand on the routes does not justify leasing additional aircraft, given the associated costs and the capacity available with the selected aircraft types.

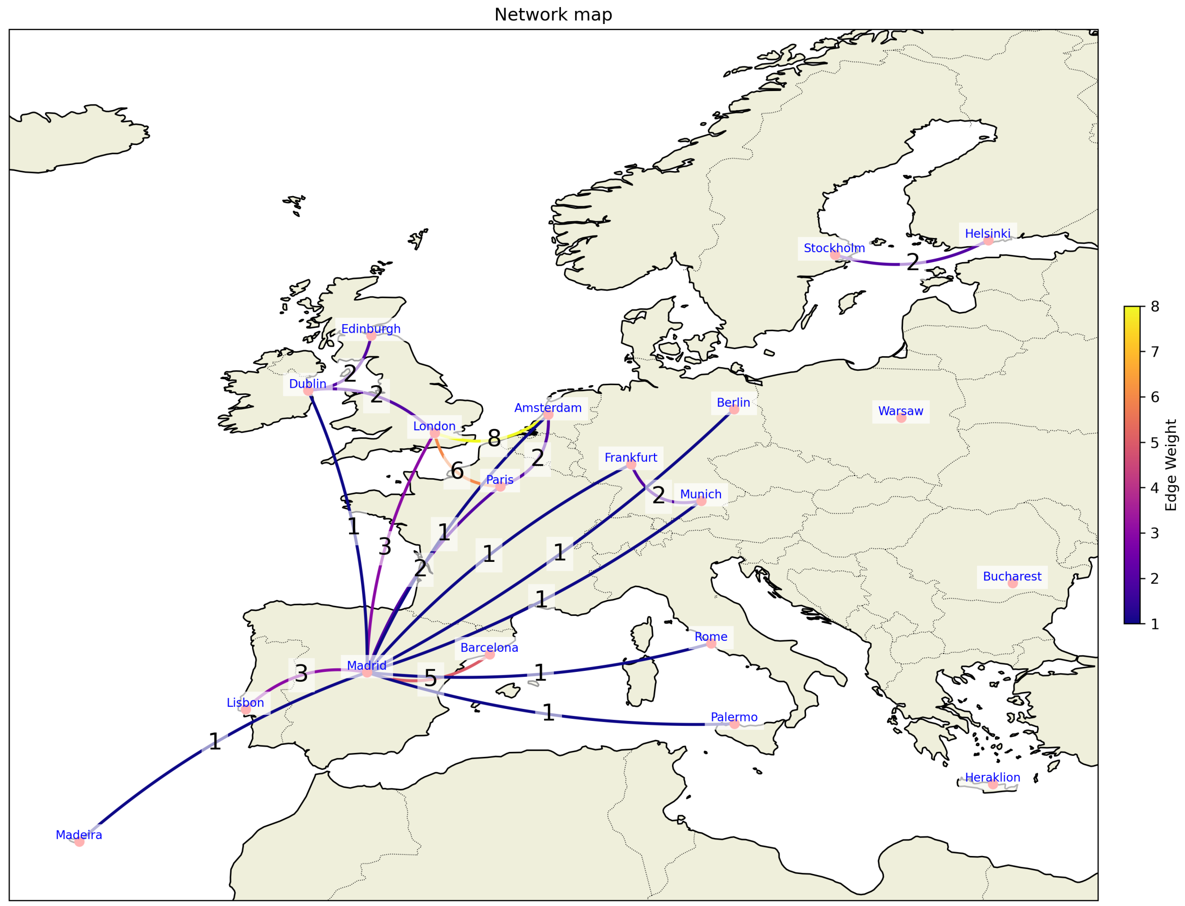

Figure 2.1: Network Map

Table 2.3: Aircraft Types and Number of Aircrafts

Aircraft Types	Number of Aircrafts
Aircraft 1: Regional turboprop	0
Aircraft 2: Regional jet	1
Aircraft 3: Single aisle, twin engine jet	1
Aircraft 4: Twin aisle, twin engine jet	0

2.2.1. KPI's

The calculated key performance indicators (KPIs) provide valuable insight into the performance of the airline network. The revenue passenger kilometer (RPK) is 4.90 million euros, which indicates a substantial volume of revenue-generating passenger traffic, reflecting the airline's ability to attract demand across its network. The yield per RPK is €0.08, which is relatively low.

The average load factor (ALLF) is 0.73, demonstrating reasonable utilization of the capacity. This indicates there is room for optimization to increase seat occupancy, which could improve overall profitability. The KPIs collectively highlight the balance between traffic demand, pricing strategies, and operational efficiency.

Table 2.4: Key Performance Indicators (KPI's)

KPI	Value
Revenue Passenger Kilometer (RPK)	4.90 million €
Yield per RPK	0.08 € per RPK
ALLF	0.73 (73%)

3

2 - Passenger mix flow

3.1. Mathematical model

The notation used for this mathematical formulation is provided in Table 3.1. This mathematical model is a keypath formulation.

Table 3.1: Notation

Sets and indices	
L	Set of flights
P	Set of all passenger itineraries (paths)
P_p	Set of passenger itineraries (paths) with recapture from itinerary p
Parameters	
$fare_p$	Average fare for itinerary p in EUR
D_p	Daily unconstrained demand for itinerary p
Q_i	Daily unconstrained demand for itinerary p
CAP_i	Capacity on flight (leg) i
b_p^r	Recapture rate of a pax that desires itinerary p and is allocated to r
δ_i^p	1 if flight leg i belongs to path p; 0 otherwise
Decision variables	
t_p^r	Number of passengers with itinerary p but are relocated to r

Where:

$$Q_i = \sum_{p \in P} \delta_i^p \times D_p, \quad \forall i \in L \tag{3.1}$$

$$\delta_i^p = \begin{cases} 1, & \text{if flightleg } i \text{ belongs to path } p \\ 0, & \text{otherwise} \end{cases}, \forall i \in L, p \in P \tag{3.2}$$

The mathematical formulation then follows as:

$$\min \quad \sum_{p\in P}\sum_{r\in P}(fare_p - b_p^r \times fare_r) \times t_p^r \tag{3.3}$$

Subject to:

$$\sum_{p\in P}\sum_{r\in P}\delta_i^p \times t_p^r - \sum_{p\in P}\sum_{r\in P}\delta_i^p \times b_p^r \times t_p^r \geq Q_i - CAP_i, \qquad \forall i \in L \tag{3.4}$$

$$\sum_{r\in P} t_p^r \leq D_p, \quad \forall p \in P \tag{3.5}$$

$$t_p^r \geq 0, \quad \forall p, r \in P \tag{3.6}$$

This mathematical formulation shows the spilled costs of relocated passengers of the wanted itinerary to another itinerary. The unconstrained daily demand is calculated with formula 3.1. δ is decided as shown in formula 3.2. The objective function (3.3) aims to minimize the spillage costs of the airline. The first constraint (3.4) ensures that the number of passengers on a flight is never more than the number of seats on a plane. The second constraint makes sure there are not more passengers on a plane than the demand is, seen in constraint 3.5. The last constraint, constraint 3.6 secures that the number of passengers is never negative.

3.2. Column generation algorithm

For the column generation algorithm dual variables are calculated alongside the primal decision variable t_p^r. With these dual variables π_i and σ^P. In this problem σ^P is always 0. A column is added to the restricted master problem when the $slackness < 0$ (3.7). As seen in formula 3.8 the slackness is calculated using the dual variables.

$$slack < 0 \tag{3.7}$$

Where:

$$slack = \left(fare_p - \sum_{i\in p}\pi_i\right) - b_p^r \times \left(fare_r - \sum_{j\in r}\pi_j\right) - \sigma_p \tag{3.8}$$

3.3. Results

3.3.1. Initial restricted master problem

Table 3.2: Initial Restricted Objective

Optimal Costs	t_p^r values	π-values
€1 298 329	$t_0^{117} = 75$	$\pi_0 = 58$
	$t_1^{117} = 12$	$\pi_1 = 58$
	$t_2^{117} = 187$	$\pi_2 = 54$
	$t_3^{117} = 0$	$\pi_3 = 0$
	$t_4^{117} = 0$	$\pi_4 = 0$
	$t_5^{117} = 131$	$\pi_5 = 40$

The keypath formulation is implemented in a column generation algorithm. The initial results of the RMP can be seen in table 3.2. The optimal costs are the result of the objective function being €1 298 329. In the initial problem, all relocated passengers go to the fictitious itinerary. The value of t_p^r shows the number of passengers who moved from their preferred itinerary to the fictitious itinerary. The dual variable π shows the reallocation costs.

3.3.2. Results passenger mix problem

In this section the results of the whole passenger mix problem will be discussed. In the following iteration, the restricted master problem is extended with three columns which have negative slackness. After this iteration, all remaining columns have negative slackness and are included in the final iteration, which results in the optimal minimized airline cost of €1 238 582, which are the spilled costs of passengers for which no seat could be assigned. There are 6312 passengers which do not take a flight at all and are thus the spilled passengers. In table 3.3 the results of the decision variable of the first 5 itineraries can be seen. The results are rounded to 1 decimal place, because the decision variables are set to be continuous to retrieve the dual variables. In actuality, the values should be integer, since we are dealing with passenger numbers . When the value is 0, there are no passengers reallocated to a different itinerary. It can be seen that there are 75 passengers that want to be reallocated from itinerary 0 to itinerary 6. Furthermore, 142 passengers are reallocated from it. 2 to it. 6. Approximately 49 passengers are want to get moved from 2 to path 10. Lastly, 12 passengers changed their preferred itinerary from path 1 to path 2. When the optimal value is 0, there are no re-allocations.

Table 3.3: Results of Passenger Mix Problem related to first 5 itineraries

	Value
Optimal airline cost (spilled cost) (EUR)	€1 238,582
Total number of spilled passengers	6312
Optimal value for t_p^r	
From path 0 to path 1	0.0
From path 0 to path 6	75.0
From path 0 to path 7	0.0
From path 0 to path 11	0.0
From path 1 to path 2	12.0
From path 1 to path 4	0.0
From path 2 to path 4	0.0
From path 2 to path 6	142.2
From path 2 to path 8	0.0
From path 2 to path 10	49.3
From path 2 to path 11	0.0
from path 3 to path 8	0.0
from path 3 to path 9	0.0
from path 4 to path 1	0.0
from path 4 to path 2	0.0
from path 4 to path 5	0.0
from path 4 to path 7	0.0
from path 4 to path 8	0.0
Optimal value for π_i	
Flight 0	58
Flight 1	58
Flight 2	54
Flight 3	0
Flight 4	0
Flight 5	40

In table 3.4 there is a comparison to optimizing the problem with and without the column generation algorithm. When not using the column generation algorithm it uses 13689 columns to solve the problem. The official runtime is 0.03 seconds, however, in reality it took a lot longer. The column generation begins in the first iteration with 118 columns. The runtime is 0.01s. In the second iteration 3 columns are added. The last iteration uses 283 columns with 0.0s runtime. As you can see the column generation algorithm uses way less columns than the optimization without this algorithm.

Table 3.4: Comparison of Columns and Runtime with and without CGA

	Without CGA	**With CGA**
Columns	13689	It 1: 118 It 2: 121 It 3: 283
Runtime (s)	0.03	It 1: 0.01 It 2: 0.01 It 3: 0.00 Total: 0.02

4 Appendix

4.1. Division of workload

Table 4.1: Task division

Name	Mathematical modelling	Programming	Reporting
Kristian Terlien (5169941)	33.33%	33.33%	33.33%
Naomi Cornelissen (5105765)	33.33%	33.33%	33.33%
Quirine Japikse (5040930)	33.33%	33.33%	33.33%