AE4423
Assignment 1

Introduction of Central Connect airways
Group 26

Students:	Viranca Balsingh	4554000
	Edward Hyde	4285174
	Daan Kerstjens	4299418
Lecturer:	Dr.ir. B.F. Lopes dos Santos	
Teacher Assistants:	Jacob Trepat Borecka	
	Sidharatha Thakur	
Date:	December 18, 2020	

Compiled and edited at Delft University of Technology, Delft, Netherlands

TU Delft

Contents

Nomenclature

Symbols problem 1

Symbol	Description
a_{ij}	Range constraint compliance matrix.
AC^k	Number of aicraft of type k in the fleet.
b_{ij}	Runway constraint compliance matrix.
$b_{1,2,3}$	Calibration parameters representing the population (1), GDP (2) and distance (3).
BT	Block time of aircraft.
$CASK$	Operating cost per available seat kilometer.
c_{ij}	Pair quality for the assignment of student i to house j.
C_T, C_T, C_F	Fixed, time dependant and fuel operational costs.
d_{ij}	Distance for travelling from airport i to airport j.
f	The fuel cost in [MU/gallon].
$g_{i,j}$	Hub presence indicator
GDP_i	The GDP for airports i and j in [MU].
I	Index set for departure airports 1, 2, ..., |I|
J	Index set for arrival airports 1, 2, ..., |J|
k	Index set for aircraft types 0, 1, 2, 3
κ	Calibration parameter representing the market size.
L^k	Lease cost of aicraft type k.
LF	Load factor of aircraft.
LTO	Landing and take-off time, i.e. Turn around time.
$pop_{i,j}$	The population for airports i and j per 1000 inhabitants.
q_{ij}, D_{ij}	Amount of demand for travelling from airport i to airport j.
s^k	Number of seats in aicraft type k.
sp	Speed or velocity of aicraft.
x_{ij}	Number of passengers flying from airport i to airport j.
w_{ij}	Number of transfer passengers flying from airport i to airport j.
z_{ij}	Number of flights for each aircraft type travelling from airport i to airport j.

Symbols problem 2

Symbol	Description
A	Airport
BP	(de)Briefing period, always 25 minutes each
DH_p	Duty Time in hours of pairing p
DT_f	Duty Time in hours of flight f
F	Set containing all flights
FH_f	Flight hours of flight f
FS	Fixed salary per day
HS_A	Hotel Costs per room at airport A, if A is crewbase $HS_A = 0$
P	Set containing all pairings
P_1	Set containing all pairings consisting of a single flight
c_p	Cost of using pair p
p	Pairing
x_p	1 if pairing p is used, 0 otherwise
δ_f^p	1 if pairing p contains flight f, 0 otherwise
π_f	Cost associated with flight f in pricing problem

1 Introduction

Central Connect is a new airline that will start operating flights within the European continent, with its main hub at Paris Charles de Gaulle Airport (LFPG). This report focuses on 19 European destinations that are available to start operating flights on. Due to the high level of competition and costs in the industry, it is important to make an efficient plan in order to safeguard and increase the profitability of Central Connect.

Therefore, the goal of this report is to develop a network and fleet plan and to find an optimal solution to the crew pairing. To match the market demand, Central Connect needs to forecast reliable demand figures to build upon a profitable network and fleet. The network and fleet plan itself, requires mathematical models in order to find an optimal solution. The last main hurdle before commencing operations is to solve the crew scheduling and pairing of the crew members. The problem is subject to various costs, duty periods and crew operating from different bases whilst complying with the ASA agreements at the same time. This problem is solved via the column generation (CG) algorithm, introduced in the lectures [2].

This report is structured in three chapters, the first chapter being this introduction. In the second chapter, the demand for the year of introduction is forecasted, to which the network is developed and determined how many aircraft should be leased. Lastly, the crew pairing problem is solved and presented in the third chapter.

2
Problem 1

In this chapter a network and fleet plan is developed for Central Connect. Since Central Connect only has outdated demand data from 2015, first the demand for 2020 needs to be forecasted. Equipped with up to date demand data, the network and fleet plan is generated via the use of a mathematical model that is optimized via Gurobi [4]

2.1. Demand Forecast

This section covers the forecasting of the demand in 2020. Besides the outdated demand data from 2015, also the latest GDP and population data stems from 2018. The forecasting of the demand is done by calibrating the widely used gravity model in long-haul transport. The linearization and calibration of this model is done in subsection 2.1.1. Followed by estimating the population and GDP data in 2020, by linear extrapolation. Lastly, combined with the calibrated gravity model and the estimated data the demand for 2020 is forecasted. All these calculations are done in the `1_Problem 1A.py` script.

2.1.1. Gravity Model Calibration

As mentioned before, a gravity model is used to forecast the demand for 2020. The gravity model for this assignment is introduced in Equation 2.1

$$D_{ij} = \kappa \frac{(pop_i pop_j)^{b1} (GDP_i GDP_j)^{b2}}{(f \cdot d_{ij})^{b3}} \tag{2.1}$$

In order to calibrate the model it first needs to be linearized. The linearization is done by applying natural logarithms. Taking the logarithm on both sides and using the properties of logarithms, results in the log-log model formulation stated in Equation 2.2

$$\begin{aligned} \ln D_{ij} &= \ln \kappa + \ln (pop_i pop_j)^{b1} + \ln (GDP_i GDP_j)^{b2} - \ln (f \cdot d_{ij})^{b3} \\ &= \ln \kappa + b1 \ln (pop_i pop_j) + b2 \ln (GDP_i GDP_j) - b3 \ln (f \cdot d_{ij}) \end{aligned} \tag{2.2}$$

With the linearized model, an ordinary least squares regression (OLS) is applied to obtain the best fit. However, before the OLS regression can be done, the distances between airports need to be determined. The distances are calculated with the use of the haversine formula[1], and by retrieving the latitudes and longitudes from `Assignment1_Problem1_Datasheets.xlsx`. A selection of distances determined are shown in Table 2.1. These distances are verified rather straightforward, by consulting a web mapping service.

Table 2.1: A selection of the distances determined with the use of the haversine formula[1]

Great circle distances [km]					
	EGLL	LFPG	EHAM	EDDF	LEMF
EGLL (London)	0	347.17	370.45	654.76	1245.95
LFPG (Paris)	347.17	0	398.27	449.01	1064.51
EHAM (Amsterdam)	370.45	398.27	0	366.58	1460.67
EDDF (Frankfurt)	654.76	449.01	366.58	0	1422.05
LEMF (Madrid)	1245.95	1064.51	1460.67	1422.05	0

Now the calibration parameter can be determined through OLS regression. The calibration parameters are determined by minimizing the residual values. To be exact, it minimises the distances between the data and the curve fit. Before doing the OLS regression it is important to realise over what dataset the regression is executed. The dataset covers 20 airports for the population and 20 countries for the GDP. Therefore the total amount of observations that need to be considered for the OLS regression is determined by Equation 2.3.

$$observations = i \times (i-1) = 20 \times 19 = 380 \tag{2.3}$$

That is way the data needs to be adjusted, because for both demand and the distances there is a diagonal with zeros. This diagonal of zeros due to symmetry of the data, would cause the regression to be biased. Then, all known data from 2015 data is put into the OLS regression according to the linearization of Equation 2.2. From Table 2.2, the values of the estimated calibration parameters are given. The calibration parameters are verified by replicating the demand data already available from 2015. Here it is observed that the estimation of demand in 2015 is in accordance with the actual demand data from 2015. Moreover, it is observed that the coefficient of determination, R^2 and the adjusted R^2 are almost both 0.8. Which indicates that the fit could be improved, due to biases being present because of log-linearized OLS regression. In most cases therefore, an estimation using a Poisson pseudo-maximum likelihood (PPML) is chosen[5]. Estimating in its multiplicative form reduces the biases and therefore the goodness of the fit.

2.1.2. Population and GDP Forecast

The last step that is required before the demand in the year 2020 can be forecasted, is to calculate the population and the GDP by assuming linear variation. Regarding the population and GDP, there is data available from 2015 and 2018. This data from 2015 and 2018 is extrapolated with the assumption of linear variation, which gives Equation 2.4 for the population

$$pop_{i,2020} = (2020-2018)\frac{pop_{i,2018} - pop_{i,2015}}{(2018-2015)} + pop_{i,2018} \tag{2.4}$$

A similar formula is used to estimate the GDP for 2020.

Table 2.2: Output of the OLS regression

OLS Regression Results			
Dep. Variable	Demand	R^2	0.801
Observations	380	Adjusted R^2	0.800
Coefficient		Value	
κ		1.4989e-6	
b_1		0.5252	
b_2		0.6036	
b_3		1.3273	

Table 2.3: A selection of the forecasted demand in 2020

Demand per week in 2020					
	EGLL	LFPG	EHAM	EDDF	LEMF
EGLL	0	1350	882	378	241
LFPG	1350	0	364	283	135
EHAM	882	364	0	264	63
EDDF	378	283	264	0	60
LEMF	241	135	63	60	0

2.1.3. Demand Forecast

With all the necessary data being available now, the demand in 2020 can be forecasted. By putting all calculated data from 2020 and the calibration parameters in the gravity model of Equation 2.1. A sample of the forecasted demand is shown in Table 2.3 [2].

2.2. Network & Fleet Development

In this section, the development of the network & fleet model is described. To begin, several assumptions made in modelling the task are discussed in subsection 2.2.1 . Afterwards, a mathematical representation of the intended model is drawn in subsection 2.2.2. Then, the various variables and parameters are discussed. Some of these are found during the forecasting of the demand, others are calculated in this subsection. Next, the results are shown using key performance indicators (KPIs). Finally, the results are analysed and discussed. Possible shortcomings of the model are put forward and light is shone on patterns in the outcome. All these calculations are done in the `1_Problem 1B.py` script.

2.2.1. Assumptions

The model build in this chapter is a representation of a real-world airline. However, some simplifications are made to constrain the scope of the case that has to be modelled. The assumptions are listed and described below. There are also obvious constraints like the number of airports and the number of aircraft types.

- **Demand is static.** In the real world, demand changes over time in real-time. In this report, the demand is considered to be constant over time. Therefore the time dimension can be neglected here, and we only optimize for one set of flight demands. On the contrary, one would want to optimize the problem continuously with the changing demands to account for this. Moreover, the demand is considered static to changes in for example flight frequency and direct versus connecting flights.

- **No competition.** Similarly to the previous assumption, the demand is again considered to be constant. This time however, it remains constant regardless of the existence of competition. In other words, competition is disregarded in the model. To account for this, one would again measure the demand continuously and optimize the model accordingly.

- **No passenger's choice.** Another assumption is the inability of the passengers to have a preference for a certain flight type or route. In reality, passengers might, e.g., wish only to fly directly to their destination. In reality this could be accounted for using the continuous demand, as well as prediction from empirical data analysis.

- **Possible aircraft routing discontinuity.** Another thing that is disregarded is the feasibility of the found aircraft routing. It could for example be the case that flight routes emerge that are distant from the rest of the network and in turn, these routes are not connected on different places in the network.

- **Single scenario for a static future.** When the model is optimized, it is based on a snapshot in time and predicts an outcome again for a single snapshot in time. This neglects the uncertainty that is bound to the future of the network. An example of this is to account for seasonality.

- **No unforeseen costs.** In reality, it is common to account for scheduled and unscheduled maintenance using risk analysis methods and perhaps set aside a maintenance budget or spread it on to the consumers. In this report, the maintenance aspect is neglected and only fixed operating, time-based, fuel and lease costs are considered.

- **No multi-hub network.** The network will only deal with a single hub, in this case Paris. In reality, large airlines could have multiple or shared hubs. This would enable to have lower operating cost at all these hubs. Also, direct hub transfers could take place at two nodes in the network, which would have an impact on the dynamics of the network in general.

- **No multi-transfer flight routes.** Passengers that are on a transfer route, always pass through the hub. There are no transfer flights through other airports. In turn, the maximum number of flights from departure to arrival airport of a passenger is two. In reality, a passenger might take three or more flights, passing through multiple airports in the process.

2.2.2. Mathematical Model

This section elaborates upon the mathematical model that is at the basis of the network & fleet model. Starting off with the objective of maximizing profit, this is captivated in Equation 2.5, the objective function. The profit is computed by computing the income from yield, where yield is estimated by Equation 2.13. From this, the operation and aircraft lease costs are subtracted. The Operating costs are estimated in three parts as shown in Equation 2.14 to 2.16. Finally, the characteristics of the various aircraft types are given in Table 2.4 [2] [1].

$$\text{Max Profit} = \sum_{i\in N}\sum_{j\in N}\left[\text{Yield}_{i,j} \times d_{ij}\left(x_{ij}+w_{ij}\right) - \sum_{k\in K}\left(CASK^k \times d_{ij} \times s^k \times z_{ij}^k\right)\right] - \sum_{k\in K}\left(L^k \times AC^k\right) \tag{2.5}$$

Equation 2.6 to 2.12 represent the constraints of the optimisation task. Constraint 1 in Equation 2.6 ensures that all passenger flow, transfer or not transfer, is smaller than the demand of those flows. In turn, this ensures that no flow is left at an airport. Equation 2.7 is an extension to this, which ensures that transfer passengers only occur if the arrival or destination airport is not the hub.

$$x_{ij} + w_{ij} \leq q_{ij} \quad, \forall i, j \in N \tag{2.6}$$

$$w_{ij} \leq q_{ij} \times g_i \times g_j, \forall i, j \in N \tag{2.7}$$

Equation 2.8 shows the second formal constraint, and ensures that the flow between airports is bounded to the capacity of the aircraft flying in between those airports. For this, the average load factor *LF* is estimated to be 80%.

$$x_{ij} + \sum_{m\in N} w_{im} \times \left(1-g_j\right) + \sum_{m\in N} w_{mj} \times \left(1-g_i\right) \leq \sum_{k\in K} z_{ij}^k \times s^k \times LF \quad, \forall i, j \in N \tag{2.8}$$

Next, a constraint is made to ensure the incoming and outgoing flight at each airport are in balance. This is important since aircraft are not supposed to remain at an airport. Moreover, it should not be possible to send out more aircraft than that come in. This is shown in Equation 2.9.

$$\sum_{j\in N} z_{ij}^k = \sum_{j\in N} z_{ji}^k, \forall i \in N, k \in \boldsymbol{K} \tag{2.9}$$

Equation 2.10 then shows constraint four which limits the use of each aircraft in the fleet to the block time of each aircraft. For this, an average of ten hours per day is taken, weekly. Something to note is the landing and take-off time, i.e. the turn around time, which is taken to be 50% larger when flying to the hub.

$$\sum_{i\in N}\sum_{j\in N}\left(\frac{d_{ij}}{sp^k} + LTO\right) \times z_{ij}^k \leq BT^k \times AC^k, k \in \boldsymbol{K} \tag{2.10}$$

The two Equations below, 2.11 and 2.12, then make sure that the range of the aircraft is not exceeded and that the aircraft only go to airports that have long enough runways, respectively.

$$z_{ij}^k \leq a_{ij}^k \quad \rightarrow a_{ij}^k = \begin{cases} 10000 & \text{if } d_{ij} \leq R^k \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

$$z_{ij}^k \leq b_{ij}^k \quad \rightarrow b_{ij}^k = \begin{cases} 10000 & \text{if } RW_{ij} \leq RW_{req}^k \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

Revenue-Passenger-Kilometer (RPK)

The yield that is called in the objective function in Equation 2.5, is computed by Equation 2.13 which is dependent on the distance of the flight from airport i to airport j [1].

$$Yield_{i,j} = 5.9 \cdot d_{ij}^{-0.76} + 0.043 \tag{2.13}$$

Operating costs

The operating costs are split up in three parts. First of all, the fixed operating costs C_X^k which can be found in Table 2.4 and depend on the aircraft type, this accounts for landing costs, parking fees and fixed fuel costs. Secondly, the time based costs, estimated by Equation 2.14. This accounts for time-dependent operating costs, e.g. cabin and flight crew wages. Thirdly, the distance depending fuel costs are expressed in 2.15. These various costs are finally summed up to form the total operating cost in Equation 2.16. Important to mention is that the operating costs for all flights departing or arriving at the hub are considered 30% lower due to economies of scale [1].

$$C_{T_{ij}}^k = c_T^k \frac{d_{ij}}{V^k} \tag{2.14}$$

$$C_{F_{ij}}^k = \frac{c_F^k \cdot f}{1.5} d_{ij} \tag{2.15}$$

$$C_{ij}^k = C_X^k + C_{T_{ij}}^k + C_{F_{ij}}^k = CASK^k \times d_{ij} \times s^k \tag{2.16}$$

Aircraft types
Finally, the parameter K represents the subset of aircraft in the fleet. This fleet is in the end chosen by the optimizer, which has the ability to choose from the four aircraft types indicated in Table 2.4. Moreover, it is allowed to choose multiple aircraft of the same type, of course bounded by the previously mentioned constraints.

Table 2.4: Aircraft type characteristics [1]

Aircraft type	**Aircraft 1:** Regional turboprop	**Aircraft 2:** Regional jet	**Aircraft 3:** Single aisle twin engine jet	**Aircraft 4:** Twin aisle, twin engine jet
Aircraft characteristics				
Speed [km/h]	550	820	850	870
Seats	45	70	150	320
Average TAT [mins]	25	35	45	60
Maximum range [km]	1,500	3,300	6,300	12,000
Runway required [m]	1,400	1,600	1,800	2,600
Cost				
Weekly lease cost [€]	15,000	34,000	80,000	190,000
Fixed operating cost C_x [€]	300	600	1250	2000
Time cost parameter C_T [€/hr.]	750	775	1400	2800
Fuel cost parameter C_f	1.0	2.0	3.75	9.0

2.2.3. Model Performance

Now that the objective function and all the constraints have been defined, the network and fleet model can be modelled and solved numerically. Using the Gurobi-Python interface, the optimal solution was found to be a profit of 13.2 thousand monetary units (MU). The models stops further optimization as soon as the difference between optimization steps is smaller than 0.5% [4].

Furthermore, the fleet that the model converges to consists of three type 1 aircraft: Regional turboprops, a single type 2 aircraft: Regional jet, and finally four type 3 aircraft: Single aisle twin-engine jets. This is also indicated in Figure 2.1. Afterwards, the model returns the frequency of each aircraft type to fly between every pair of airports. In Figure 2.2, the non-trivial flight path frequencies are shown visually.

Finally, some indicators are drawn from the optimal solution. First of all is the seat saturation, which is a metric that shows per aircraft type the number of filled seats as a percentage of the number of available seats with the current network and fleet. Secondly, the utilization of the aircraft is computed, which is the usage time of the aircraft as a percentage of their available block time. Next, the number of flights and the number of leased aircraft are shown in Figure 2.1, this comprises the KPIs.

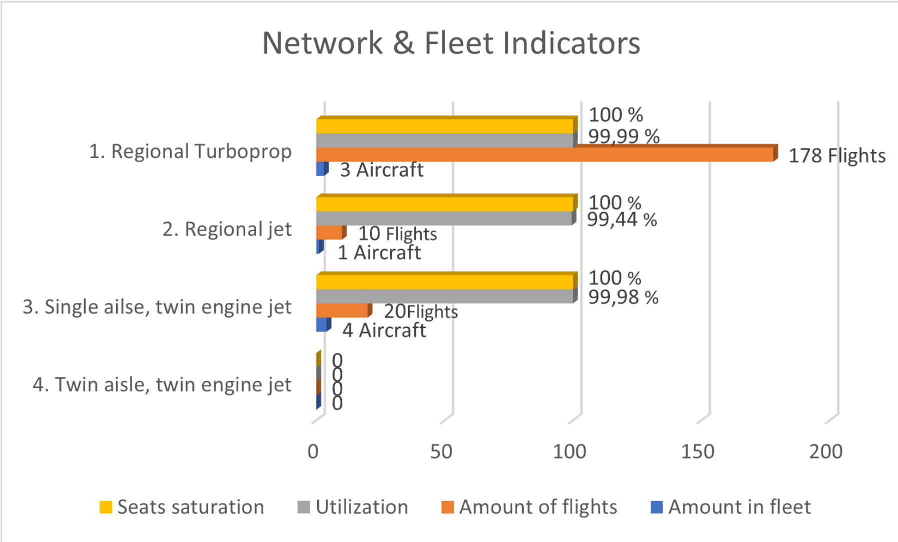

Figure 2.1: Key Performance Indicators

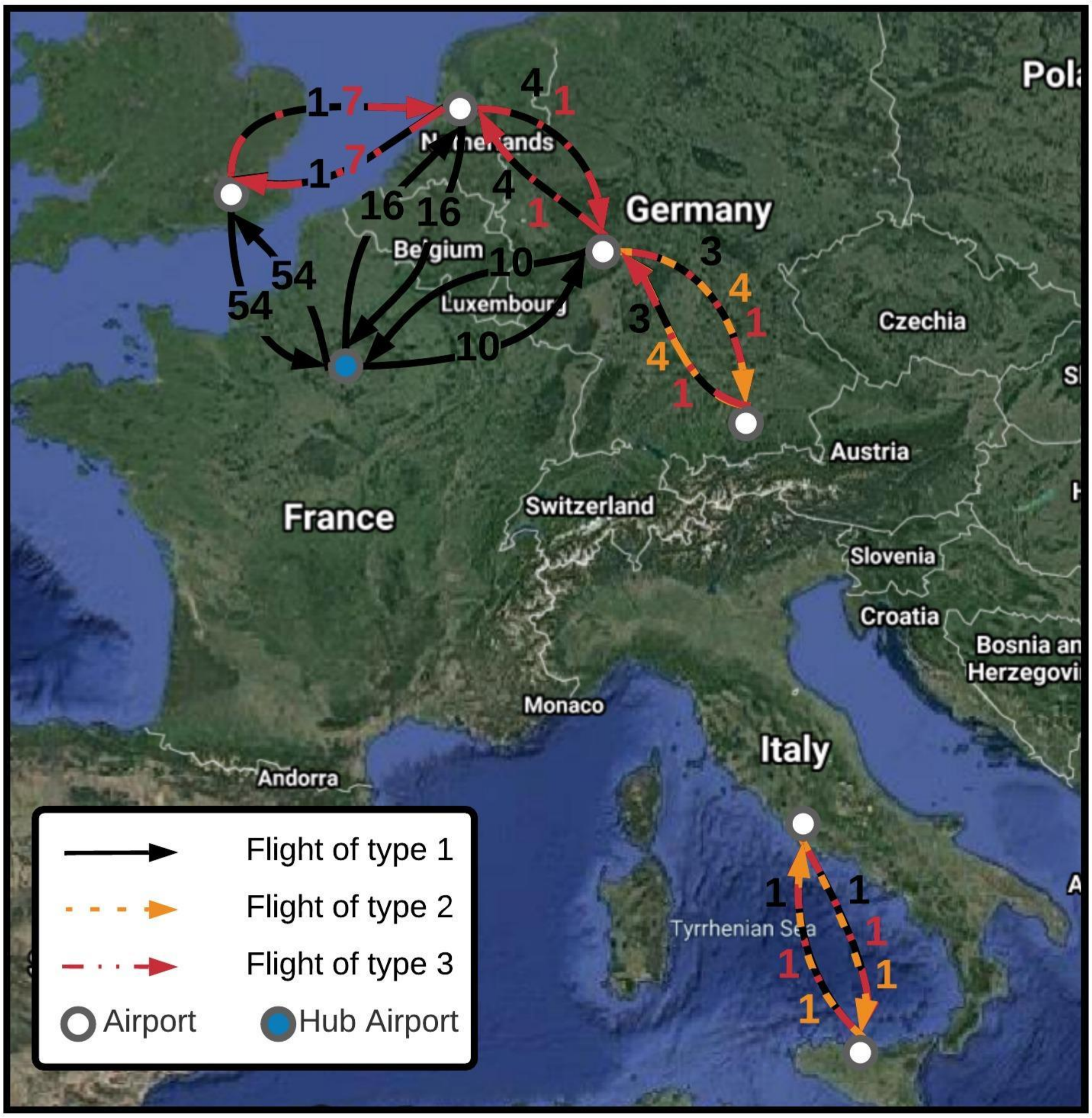

Figure 2.2: Flight paths and frequencies [3]

2.2.4. Model Discussion

This subsection will analyze the results argues whether the findings are sensible or not, and why so. Moreover, the findings will be linked to the simplifications/assumptions made during the mathematical modelling of the task at hand.

Looking again at Figure 2.2, a couple of things are noticeable. First of all, type 1 aircraft are significantly more common in flights that are connected to the hub directly. This is sensible as these small aircraft types are generally more profitable on shorter flights. In shorter flights, landing and take-off costs play a larger roll in the total profit to be made from such flights. Therefore, having these short flights to benefit most from the discounts of the hub airport is sensible to maximize profit. The same line of argumentation can be considered when the other aircraft types are observed. Aircraft type 3 is used most often for flights not reaching the hub, likely the benefit of fitting more passengers is larger than the cheaper operating costs of smaller jets in those cases as there is no hub-discount.

Another thing to note is that the current solution shows no usage of the fourth aircraft type. It seems that this aircraft is most suitable for long haul flights, which is something that is also seen in practice from real world examples. To further test this, the model would have to be extended with more distant airports, as well as with a new assessment of the assumptions and constraints that are important in such a scenario.

Moreover, it is visible that the network is divided into two subnetworks. Where the airports in Italy, Pallermo and Rome are separated from the rest of the involved airports. In practice this would not be optimal as it is better to be able to fly back any aircraft to the hub, via flights that are part of the network. Now, when an aircraft in Italy requires maintenance, it has to first fly back to the hub. This extra flight is obviously detrimental for profit and should be avoided. To improve this, a constraint should be included which ensures that all airports that are non-trivial, will have a direct or indirect connection to the hub.

Finally, even though the number of aircraft of each type in the fleet does not deviate much, the amount of flights is significantly larger for the type 1 aircraft than for the other two types. This is again sensible as the type 1 aircraft is more profitable for short flights as it has fewer seats and a smaller turn around time, e.g. a KLM Cityhopper. Additionally, as the flights are shorter, more flights fit in the assigned block time of the aircraft.

3
Problem 2

Having developed a network and fleet plan, Central Connect is almost ready to commence operations. In this chapter, a simplified version of the crew pairing problem is discussed. The crew pairing problem is introduced first, after which the assumptions that are accompanied with this simplification are discussed in subsection 3.1.1. Next, the costs that are associated with each duty period are computed in subsection 3.1.2. Then, the crew pairing problem is solved initially with duty periods that consist only of a single flight in subsection 3.1.3. Followed by the column generation algorithm to solve the crew pairing problem in subsection 3.1.4. Lastly, the results are analysed and discussed and possible limitations to the model are discussed.

3.1. Crew Pairing Problem

The duty periods were pre-generated, and already comply with both ASA and labour agreements, hence the scheduling problem reduces to a Crew Pairing Problem[2]. The Crew Pairing problem tries to solve the objective:

$$\min \sum_{p \in P} c_p \times x_p \tag{3.1}$$

Subject to

$$\sum_{p \in P} \delta_f^p \times x_p = 1, \quad \forall f \in F \tag{3.2}$$

$$x_p \in \{0, 1\}, \quad \forall p \in P \tag{3.3}$$

Since the set P of possible pairing is enormous, only flights of a single day are considered, and some assumptions, as explained in the next section, are made. First a initial feasible solution is generated in subsection 3.1.3 and than the solution is improved with column generation in subsection 3.1.4.

3.1.1. Assumptions

The model uses many assumptions to reduce computation times.

- **The crew base is the same as the departure of the first flight in each duty**. This eases the model and makes it possible to make a pairing of a single day. Because of this assumption, it is well described how many crew at each base is needed.
- **Pairings considered in optimum continuous solution will give the optimum solution in the binary problem as well.** To obtain the dual problem, the binary condition has to be relaxed. The final solution of the pricing problem therefore has (a few) non-integer values for pairings. The final solution of the crew pairing problem is calculated with these rows (in dual problem, columns in primal problem).
- **All flights have same amount of each crew type.** i.e. a captain, a first officer and three flight attendants.
- **Next duty period starts at the destination airport of the last flight in a duty period.** It is assumed all crew will have an available duty the next day starting from this airport.
- **Overnight stay has lowest cost.** It is assumed it is more profitable to have crew stay for the night than to put them on a deadhead back to crew base. This makes sense in combination with the previous assumption, i.e. the next duty period starting at the same base.
- **Every airport could be a crew base.** In reality, it might undesirable to have certain airports as crew base, either because of labour laws or the challenge of finding competent crew near that airport.
- **Crew costs are equal for every crew base.** This implies no transportation costs, and no different labour agreements are in place. In reality, crew cost may vary depending on country of residence.

3.1.2. Costs

The cost of each duty is composed of three parts:

- A fixed salary per day
- An hourly salary per duty hour. Duty hours are calculated by the duty duration of each flight, including a brief period of 25 minutes before each flight and a debrief period of 25 minutes after each flight.
- Hotel costs for crew if they do not return to their home base at the end of a duty cycle.

The cost of a single pairing $c_{p,m}$ is therefore given by:

$$c_{p,m} = FS_m + DP_m \times DH_p + HC_{A,m} \tag{3.4}$$

In which the subscript m denotes a single crew member. Hence the total cost of a duty period is given by:

$$c_p = \begin{pmatrix}1\\1\\3\end{pmatrix} \cdot \left[\begin{pmatrix}98\\35\\15\end{pmatrix} + \begin{pmatrix}120\\55\\18\end{pmatrix} \sum_{f \in p} DH_f + \begin{pmatrix}1\\1\\1\end{pmatrix} HC_A \right] \tag{3.5}$$

In this formula, the arrays represent the partitioned amounts and costs of the different crew members, with on the top row captain, centre row first officer and bottom row flight attendants. The duty time of a flight contains its flight time + 2 (de)brief periods of 25 minutes each: $DH_f = FH + 2BP$. The duty costs are calculated with `2_Problem_2_dutyCosts.py`. An example of some duty costs is given in Table 3.1.

3.1.3. Initial Crew Pairing

The initial crew pairing problem is solved to have a feasible solution to start with in the Reduced Master Plan (RMP). It has the same objective (Equation 3.1) and constraints (Equation 3.2 and 3.3) as the entire crew pairing problem, but only pairs consisting of a single flight P_1 are considered, instead of the complete set P. Only one feasible solution exists, namely

$$x_p = \begin{cases} 1, & \forall p \in P_1 \\ 0, & \text{For all other } p \end{cases} \tag{3.6}$$

resulting in a total cost of 213 543 MU. This solution was found by `2_Problem_2_InitialProblem.py` in 0.03 seconds.

3.1.4. Pricing Problem

For the column generation algorithm, we have to look at the dual problem, in this case the pricing problem. Since a MIP has no dual, the constraints need to be relaxed to remove all equal signs and binary conditions. Hence the new constraints of the primal problem are:

$$\sum_{p \in P} \delta_f^p \times x_p \geq 1, \quad \forall f \in F \tag{3.7}$$

$$x_p \geq 0, \quad \forall p \in P \tag{3.8}$$

Relaxation 3.7 will not affect the outcome, since the minimization objective still prevents values greater than one. Objective 3.1 and constraints 3.7 and 3.8 can be translated to a dual problem:

Objective:

$$\max \sum_{f \in F} 1 \times \pi_f \tag{3.9}$$

Subject to:

$$\sum_{f \in F} \delta_f^p \times \pi_f \leq c_p, \quad p \in P \tag{3.10}$$

$$\pi_f \geq 0, \quad \forall f \in F \tag{3.11}$$

The pricing problem starts with only pairings $p \in P_1$ too, but extra pairings will be added iteratively as long as the objective is decreasing. After each evaluation, the slackness is calculated with:

$$\text{slackness}_p = c_p - \sum_{f \in F} \delta_f^p \times \pi_f, \quad \forall p \in P \tag{3.12}$$

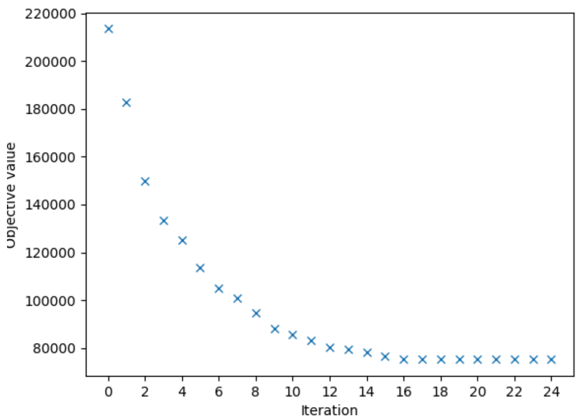

Figure 3.1: Objective value plotted against the number of iterations

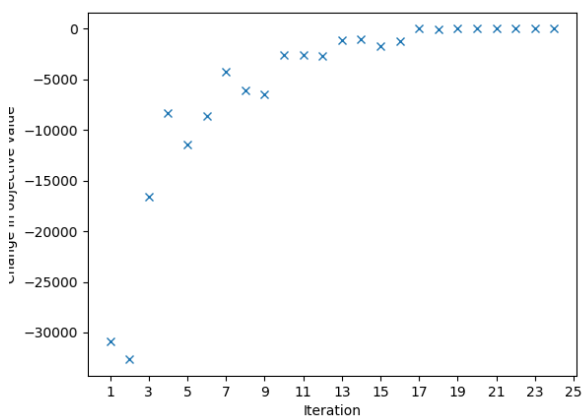

Figure 3.2: Change in the objective after each iteration. The decrease of the total cost becomes smaller after each iteration

The primal problem is optimal if the dual problem is feasible. For the pairs already considered in the evaluation, these slackness values will always be equal or greater than 0 (otherwise the solution was not feasible yet). Adding the pairing, and hence a constraint in the dual problem, with the most negative slackness, will reduce the total cost of the problem the most. Therefore, the 50 pairings with the most negative slackness will be added and the model is run again, until the change in the total cost is less than 0.001 MU for the last 6 iterations. This is done by the script `2_Problem_2_PricingProblem.py`. The KPIs of this problem are computational time, convergence of the graphs as well as the objective value.. Convergence is shown in Figure 3.1, the first few iterations have the most impact on the final solution, while the last iterations have marginal effect on the solution, asymptotically approaching the final optimum value. The change of the objective function per iteration, i.e. the cost reduction of adding the pairings, clearly decreases, as shown in Figure 3.2, and asymptotic approaching 0 (and for the last few iterations becoming zero).
The column generation iterated 24 times, which took 0.6 seconds, resulting in an objective value of 75 152.83 MU.

3.1.5. Results of Model's Key Performance Indicators

For the final result, the primal problem with the initial constraints (3.2 and 3.3) is again calculated, with the added pairings of the column generation algorithm described in the previous section. The main difference is the re-application of the binary conditions, which had to be relaxed for the dual problem. After calculating the solution, the model gives a schedule with 46 used duties servicing all 141 flights, of which the first 20 are presented in Table 3.3. The total crew cost per day running these duties is 75 152.83 MU. Of these costs, 29410.00 MU arise from hotel costs. The required crew at each base is given in Table 3.2.

3.1.6. Discussion

Since London Heathrow and Charles de Gaulle are visited most frequent by Central Connect, it is no wonder these airports need the most crew. However, these numbers are not conclusive, since the computed crew pairing can not be repeated every day with the same flight schedule: the number of utilized duty periods ending at an airport is not the same as the number of utilized pairings leaving the same airport. In reality, one can either choose to combine pairings over multiple days to ensure the return of each crew, or add a constraint that on a single day the number of duties starting at an airport need to be equal to the number of duties terminating at that airport.

$$\sum_{p\in P} x_p \times orn_p^A = \sum_{p\in P} x_p \times dst_p^A, \quad \forall A \tag{3.13}$$

with

$$orn_p^A = \begin{cases} 1 & \text{if origin of first flight in} \\ & \text{pairing } p \text{ is at airport } A \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad dst_p^A = \begin{cases} 1 & \text{if destination of last flight in} \\ & \text{pairing } p \text{ is at airport } A \\ 0 & \text{otherwise} \end{cases}$$

Table 3.1: Total duty costs of some duty periods. These costs include fixed salary, (hourly) duty pay and, if the duty period does not return to the crew's base, hotel costs

Duty Period #	Total Cost (MU)
100	1200.58
500	3053.25
750	1276.17
1200	2371.50
2000	1742.83

Table 3.2: Required crews per base. The mentioned numbers are total crews, each consisting of a captain, a first officer and 3 flight attendant, e.g. Charles de Gaulle needs 16 captains, 16 FOs and 48 FAs

Airport Code	Airport Name	Required Crew
EDDF	Frankfurt am Main	6
EGLL	London Heathrow	15
EHAM	Amsterdam Schiphol Airport	7
LFPG	Charles de Gaulle Airport	16
EIDW	Dublin Airport	1
ESSA	Stockholm Arlanda	1

Table 3.3: Amount of flights in duty period, first flight of duty period, origin of first flight, hotel costs for crew and total duty cost for the first 20 used duties by duty identifier

Duty Identifier	# Flights	First Flight	Crew Base	Hotel Costs (MU)	Total Duty Cost (MU)
154	4	CC3840	EDDF	540.00	2702.67
173	4	CC1638	EGLL	650.00	2469.17
215	5	CC3778	EHAM	1215.00	3358.58
265	5	CC2410	LFPG	705.00	3039.42
307	3	CC1638	EGLL	705.00	2237.92
315	4	CC3897	EDDF	1215.00	3473.08
344	5	CC3677	EHAM	555.00	2774.92
389	3	CC2410	LFPG	785.00	2222.50
410	4	CC3677	EHAM	785.00	2546.92
417	5	CC1210	EGLL	555.00	2832.17
477	5	CC2745	LFPG	660.00	2956.25
491	4	CC2175	LFPG	0.00	1781.00
492	5	CC3621	EHAM	555.00	2851.25
493	4	CC3897	EDDF	0.00	1819.17
542	4	CC1918	LFPG	0.00	1781.00
557	5	CC1674	EGLL	1215.00	3568.50
560	5	CC2097	LFPG	650.00	2850.83
567	4	CC1152	EGLL	660.00	2612.75
577	3	CC3840	EDDF	660.00	2116.58
582	5	CC1210	EGLL	1215.00	3454.00

The found objective is a feasible optimum if it is feasible in both the primal and the dual. However, the solution found in the dual (the pricing problem) is feasible in the relaxed primal, but not after reapplying the binary conditions. After recalculating the primal objective with the original constraints, the same value is found as in the pricing problem, hence it actually is a feasible optimum.

If assumptions on possible crew bases change, the outcome could become infeasible, i.e. if any of the airport in Table 3.2 can not be used as crew base. The objective value will also change if crew costs depend on country of residence, and this may change the optimum as well if differences are large.

Bibliography

[1] Dr.ir. B.F. Lopes dos Santos. *AE4423 Airline Planning & Optimization Assignment*. Delft University of Technology, 2020.

[2] Dr.ir. B.F. Lopes dos Santos. *AE4423 Lecture 6 - Crew Scheduling*. Delft University of Technology, 2020.

[3] Google. Google maps. `https://www.google.com/maps/@51.2756636,3.9732973,1004182m/data=!3m1!1e3`, 2020.

[4] LLC Gurobi Optimization. Gurobi optimizer reference manual, 2020. URL `http://www.gurobi.com`.

[5] Santos-Silva (J.M.C.) Tenreyro (S.). The log of gravity. *The Review of Economics and Statistics*, 88:641–658, 2006. doi: https://doi.org/10.1162/rest.88.4.641.