# Robust and Recoverable Maintenance Routing Schedules

Niklaus Eggenberg Matteo Salani<sup>∗</sup> Michel Bierlaire

January 15, 2010

#### Abstract

We present a methodology to compute more efficient airline schedules that are less sensitive to delay and can be recovered at lower cost in case of severe disruptions.

We modify an original schedule by flight re-timing with the intent of improving some structural properties of the schedule. We then apply the new schedules on different disruption scenarios and then recover the disrupted schedule with the same recovery algorithm. We show that solutions with improved structural properties better absorb delays and are more efficiently recoverable than the original schedule.

We provide computational evidence using the public data provided by the ROADEF Challenge  $2009<sup>1</sup>$ .

Keywods: Airline scheduling, Robust optimization, Disruption recovery

## 1 Introduction

In the modern society, the demand for transportation, of goods and people, is constantly increasing in terms of volume and distance. In particular, as the fastest transportation mode for mid and long distances, airline transportation develops at an impressive rate. Due to the competition between the airlines, many of them use operations research techniques to schedule their operations. This allows to keep prices low and thus attract customers while making profit. Airlines have to deal with irregular events, called *disruptions*, making the schedule unfeasible. The process of repairing a disrupted schedule is known as the recovery problem. It aims at retrieving the initial schedule as quickly as possible while minimizing the recovery costs incurred by recovery decisions (typically delaying or canceling flights).

<sup>∗</sup> corresponding author matteo.salani@epfl.ch

 $1$ http:// $\verb|challenge.roadef.org/2009/index_en.htm$ 

A major drawback of optimized schedules is that they are sensitive to perturbations. Small disruptions propagate through the whole schedule, and may have a huge impact.

The focus of this study is to implicitly consider the occurrence of future disruptions at the planing phase in order to ameliorate two properties of the schedule, namely:

- 1. the robustness: the ability of the schedule to remain feasible in the presence of small disruptions;
- 2. the recoverability: the average performance of the recovery algorithm when the schedule is disrupted.

At the planing phase, we solve the Maintenance Routing Problem (MRP), which aims at finding a feasible route for each aircraft and a deparutre time for each flight minimizing the loss of revenue as a metric which depends on the deviation from a desired schedule.

On the day of operation, the problem of recovering the planed schedule from a disrupted state is the Aircraft Recovery Problem (ARP) given the original schedule and the current disrupted state. The recovery costs for the ARP are mainly delay and cancellation costs.

The originality of the proposed algorithms is the absence of any explicit predictive model of possible disruptions for the scheduling problem. Uncertainty Features capture implicitly the uncertainty the problem is due to. An additional budget constraint ensures that the obtained solution is not too far from the original deterministic optimum, and the computational complexity is similar to the original deterministic problem.

We solve the MRP by applying the Uncertainty Feature Optimization (UFO) framework of Eggenberg et al. (2009) on a real case study and we present computational results for different MRPs using public instances of the ROADEF Challenge 2009. Recovery statistics are obtained with the recovery algorithm presented in Eggenberg et al. (forthcoming).

### 2 Literature Review

For a detailed description on the airline scheduling process, see Rosenberger et al. (2003a); for general surveys on airline scheduling and recovery problems, we refer to Clausen et al. (forthcoming), Kohl et al. (2007), Weide (2009) and Eggenberg et al. (forthcoming).

Airline Scheduling Barnhart et al. (1998a) introduce the *string* based fleeting and routing model, where a string is a sequence of connected flights between two maintenances. The problem is solved using a Column Generation scheme.This model is the reference for solving the MRP; it is used, for example, by Ageeva (2000), Rosenberger et al. (2004) and Lan et al. (2006).

Rosenberger et al. (2004) solve a robust fleet assignment problem where maximizing short cycles and hub isolation aims at improving the short cycle cancellation recovery strategy. The authors conclude that using sub-optimal solutions of the deterministic problem allow for improving a schedule's robustness.

Bian et al. (2005) study the robust airline fleet schedules for KLM, which is among the largest European airlines, showing that robustness is correlated with the number of aircrafts on ground. The presented results on eleven schedules of KLM in the year 2002 show a significant correlation between the plane on ground metric and the arrival and departure punctuality predictions.

Lan et al.  $(2006)$  propose two flight retiming models for solving the MRP. The former aims at reducing the delay propagation and the latter at reducing the missed passenger connections. In the reported results, the robust schedules allow for a reduction of about 40% of disrupted passengers and the total passenger delay is reduced by  $20\%$ .

Shebalov and Klabjan (2006) modify original crew schedules in order to maximize the move-up crews, i.e. pairings that can be swapped in operations. The main conclusion is that the trade-off between crew cost and the robustness factor is crucial: too large an investment in terms of additional crew costs to impose robustness leads to increased operational costs.

Yen and Birge (2006) describe a stochastic integer programming algorithm to solve the crew scheduling problem. Interestingly, the obtained solutions exhibit a simple but constant property: the crew tend to stay on the same plane as much as possible. The solutions show an increased average connection time between two successive flights.

Airline Recovery The literature on recovery algorithms developed mainly in the last 15 years, motivated by the growth of air traffic.

Argüello et al.  $(1997)$  and Bard et al.  $(2001)$  use a time-band model to solve the ARP. An extension of this mode is presented by Thengvall et al. (2000). They penalize the deviation from the original schedule and they allow human planners to specify preferences related to the recovery operations.

Eggenberg et al. (forthcoming) introduce the constraint specific network model for solving the general unit recovery problem, where a unit is either an aircraft, a crew member (or team) or a passenger; each unit is associated with a network encoding all feasible routes for the unit.

The literature shows that deterministic models do not lead to operationally efficient solutions. But non-deterministic models have a larger computational complexity. Remarkably, many authors conclude that more robust or recoverable solutions exhibit some improved structural properties of the solutions related to the number of aircraft on ground, the number of potential swaps (both for crew and aircraft) in the recovery phase or an increased idle time.

This motivates the use of the UFO framework of Eggenberg et al. (2009), which considers uncertainty implicitly through such features. This allows to keep the computational complexity similar to the deterministic problem.

### 3 Models and Algorithms

The global structure of both MRP and ARP algorithms is a Column Generation scheme based on the constraint-specific networks presented in Eggenberg et al. (forthcoming). As the two problems are similar, we use the same notation for both of them. Note that despite the structural similarities of the models, the MRP and ARP have different objectives, which is modeled by an appropriate cost structure. Additionally, the unit-specific constraints are modeled by a set of resources, as described in Eggenberg et al. (forthcoming).

We denote F the set of flights to be covered and P the set of available planes. S denotes the set of final states. Each of them corresponds to the expected location at the end of the scheduling/recovery period, and is characterized by an aircraft type, a location, a latest arrival time and maximal allowed resource consumption. T is the length of considered the period, which corresponds to the scheduling period for the MRP and the recovery period for the ARP. A route r is defined by the covered flights in the route, the final state and the plane. Let  $Ω$  be the set of all feasible routes r,  $x<sub>r</sub>$  the binary variable being 1 if route r is chosen in the solution and 0 otherwise, and  $c_r$  the cost of route r. Variables  $y_f$  capture flight cancellation and are 1 if flight f is canceled, incurring cost  $c_f$ , and 0 otherwise; note that for the MRP, flight cancellation is not allowed and  $c_f = \infty$ .

We define the time-space intervals  $\ell = (\mathfrak{a}, \mathfrak{t})$  to account for airport capacities. t is the index of a discretized time period (starting from

index 0) of length  $\Delta$  (typically  $\Delta = 60$  minutes),  $a \in A$  is the airport. We denote L the set of all such intervals, of cardinality  $| A | \times \left[ \frac{1}{\Delta} \right]$ . For each interval  $\ell \in L$ , the maximum number of departures is denoted by  $q_{\ell}^{\text{Dep}}$  and the maximum number of arrivals by  $q_{\ell}^{\text{Arr}}$ .

We also introduce the following set of binary coefficients:  $\mathfrak{b}_r^f$ , 1 if route  $r$  covers flight  $f \in F$ , 0 otherwise;  $b_r^s$ , 1 if route  $r$  reaches the final state  $s \in S$ , 0 otherwise;  $b_r^p$ , 1 if route r is assigned to plane  $p \in P$ , 0 otherwise;  $b_r^{\text{Dep}, \ell}$ , 1 if there is a flight in route r departing within time-space interval  $\ell \in L$ , 0 otherwise;  $b_r^{\text{Arr}, \ell}$ , 1 if there is a flight in route r arriving within time-space interval  $\ell \in L$ , 0 otherwise.

With this notation, the Master Problem (MP) of both the MRP and the ARP is the following integer linear program:

$$
\min z_{MP} = \sum_{r \in \Omega} c_r x_r + \sum_{f \in F} c_f y_f \tag{1}
$$

$$
\sum_{r \in \Omega} b_r^f x_r + y_f = 1 \qquad \qquad \forall f \in F \tag{2}
$$

$$
\sum_{r \in \Omega} b_r^s x_r = 1 \qquad \qquad \forall s \in S \tag{3}
$$

$$
\sum_{r \in \Omega} b_r^p x_r \le 1 \hspace{2.2cm} \forall p \in P \hspace{1.2cm} (4)
$$

$$
\sum_{r \in \Omega} b_r^{Dep,\ell} x_r \leq q_\ell^{Dep} \qquad \qquad \forall \ell \in L \tag{5}
$$

$$
\sum_{r \in \Omega} b_r^{\text{Arr}, \ell} x_r \le q_\ell^{\text{Arr}} \qquad \qquad \forall \ell \in L \tag{6}
$$

$$
x_r \in \{0, 1\} \qquad \forall r \in \Omega \qquad (7)
$$

$$
\mathbf{y}_{f} \in \{0, 1\} \qquad \forall f \in F \qquad (8)
$$

Objective (1) minimizes total costs. Constraints (2) ensure that each flight is covered by exactly one route  $r \in \Omega$ . Constraints (3) ensure that each final state is reached by a plane and constraints (4) ensure each aircraft is assigned to at most one route. Finally, constraints (5) and (6) ensure the departure and arrival capacities of the airports are satisfied, and constraints (7) ensure integrality of the variables.

The Column Generation process combines solving the linear relaxation of (MP) and branching to find an integer solution. The pricing problem aims at finding new feasible columns improving the current (partial) solution of the linear relaxation. It is solved as a Resource-Constrained Elementary Shortest Path Problem (RCESPP) on the constraint-specific networks. We use the dynamic programming algorithm described by Righini and Salani (2006), which is a bidirectional label setting algorithm. The algorithm creates labels, corresponding to partial paths, at each node of the constraint-specific network; dominated labels, that are proved to lead to sub-optimal paths, are discarded.

The main difference between the MRP and the ARP algorithms is the specification of the constraint specific networks and its cost structure. For the MRP, all flights are potentially feasible for an aircraft, unless the aircraft is technically not able to cover them. However, using a different aircraft than desired for a given flight may incur a loss of revenue. Such costs, in addition to retiming costs, are captured independently for each aircraft in its associated constraint-specific network and determine the costs of a route. In the ARP, the cost of a route is the sum of delay costs; the feasible flights and feasible final states are usually restricted those originally assigned to aircrafts of the same fleet type.

#### 3.1 Uncertainty Feature Optimization

The problem (1)-(8) is a deterministic model. As discussed in Eggenberg et al. (2009), using deterministic models for problems due to imperfect information leads to unstable solutions, i.e. sensitive to data variations. The MRP is clearly prone to noisy data; the nature of the noise is, however, difficult to capture due to the many factors influencing an airline's schedule: meteorological changes, economical factors such as the price of fuel, human factors such as crew illness, crew strikes, political manifestations, etc. Deriving an explicit model of the uncertainty through the characterization of an uncertainty set, is thus a difficult problem itself. As the MRP is already an NP-hard problem in its deterministic form, it is extremely hard to solve general MRP problems accounting for an uncertainty set. Finally, as shown in Eggenberg et al. (2009), solutions computed with a model involving an explicit uncertainty set are sensitive to errors in the uncertainty characterization.

An Uncertainty Feature (UF) is a structural property of a solution that is known to perform well for a general type of noise: for example, an increased idle time is known to allow for more delay absorption; increasing idle time thus improves the robustness of a solution against delays of any form; additionally, no specification of the delays is required.

When selecting UFs, we both have to consider their potential in

terms of robustness and recoverability and in terms of the implications on the algorithm. In order to preserve the column generation structure, the UFs must be formulated linearly.

#### 3.2 UFO reformulation of the MRP

The initial objective of the MRP is to find a feasible solution for the plane routing as close as possible to the input schedule; the cost  $c_r$  of route  $r \in \Omega$  is the total number of minutes the flights of route r deviate from their desired departure times, which has to be minimized.

In the framework described by Eggenberg et al. (2009), the initial objective  $\sum_{r \in \Omega} c_r x_r$  is relaxed as the following budget constraint:

$$
\sum_{r\in\Omega}c_rx_r\leq (1+\rho)z^*_{\text{MRP}},
$$

where  $\rho$  is the *budget ratio*. However, the optimal solution for the MRP is  $z^*_{MRP} = 0$ , i.e. all flights are scheduled as desired and the relative budget constraint does not allow for any change in the schedule.

We therefore use an absolute budget, with a constant C. We get the following formulation:

$$
\max z_{\text{UFO}} = \mu(\mathbf{x}) \tag{9}
$$

$$
\sum_{r \in \Omega} c_r x_r \le C \tag{10}
$$

$$
(2) - (8) \tag{11}
$$

The budget C is an upper bound on the total deviation (in time units) between original and new schedule.

Note that the additional budget constraint (10) changes the definition of the reduced cost of a column: the cost  $c_r$  is multiplied by the dual multiplier of the budget constraint in the reduced cost formulation. The structure of the pricing problem highly depends on the chosen UF  $\mu(\mathbf{x})$ , which we present, along with the implications for the pricing problem, in the next section.

# 4 Uncertainty Features for the Maintenance Routing Problem

The UFs are designed based on what practitioners do in reality: increasing idle time, which allows for delay absorption, increasing the number of plane crossings, which allows for more plane swaps in the ARP and increasing the connecting passenger's connection time. We postulate that solutions with higher values for these properties are featuring more robustness and recoverability.

#### 4.1 The IT and MIT models

The idle time of a single route is

$$
\mu_{IT}({\bf x})=\sum_{r\in\Omega}\delta_r x_r,
$$

where  $\delta_r$  is the total idle time on route  $r$ 

Using  $\mu_{\text{IT}}$  leads to a linear UFO formulation, and the structure of the pricing problem is not changed: it remains an RCESPP where the total idle time corresponds to the cost  $\delta_r$  of the column.

 $\mu_{\text{IT}}$  accounts for the total idle time. An alternative is to maximize the minimal idle time in order to get smaller but more uniformly distributed buffer time windows, i.e. use

$$
\zeta=-\min_{r\in\Omega}\delta^{\tt{min}}_r x_r,
$$

where  $\delta_r^{\min}$  is the minimal idle time in route r. This UF is however no longer linear but can be reformulated as

$$
\max - \zeta
$$
  
s.t.  $\zeta \ge \delta_r^{\min} x_r$   
 $(10) - (11)$   
  $\forall r \in \Omega$ 

However there is an exponential number of variables and constraints (at least  $| \Omega |$ ), which is not affordable for Column Generation.

Therefore we maximize the sum of the minimal idle times of each route with the following UF:

$$
\mu_{\text{MIT}}(\mathbf{x}) = \sum_{r \in \Omega} \delta_r^{\text{min}} x_r.
$$

The resulting UFO formulation is the same than for  $\mu_{\text{IT}}$ , except

we use  $\delta_r^{\min}$  instead of  $\delta_r$ . For the pricing the structure remains an RCESPP. The algorithm must however consider adapted label domination criteria. Unlike the total idle time, which is a cumulative metric during the label extension phase, the minimal idle time is decreasing in a non-homogeneous way. In order to compare labels and discard suboptimal ones, the partial reduced cost must contain a partial value of the minimal idle time that is comparable for different labels. This is the case when the minimal idle time is computed up to the end of the end of the last activity.

### 4.2 The CROSS model

The CROSS model captures the number of plane crossings, allowing for more swapping possibilities to facilitate recovery. It is not associated to a single route and thus unmanageable in the current CG scheme.

To address this issue, we introduce the concept of meeting points: we create a constraint for each airport for a discretized number of time intervals. We denote such a meeting point by the pair  $m = (a, t)$ , corresponding to the meeting point at airport a and time interval t; the number  $\Delta$  of time intervals is a fixed parameter, and M is the set of all meeting points, i.e.  $M = \{(\mathfrak{a}, \mathfrak{t}) \mid \mathfrak{a} \in A, \mathfrak{t} = 0, \cdots, \Delta\}$ . The number | M | of meeting point constraints is pseudo-polynomial (number of airports times number of time intervals).

We denote by  $b_r^m$  the binary coefficient being 1 if route r visits meeting point  $m \in M$  and 0 otherwise. We then include the following set of constraints:

$$
\sum_{r \in \Omega} b_r^m x_r - y_m \ge 0 \qquad \forall m \in M. \tag{12}
$$

The UF corresponding to the plane crossing maximization is

$$
\mu_{CROSS}(\mathbf{x}) = \sum_{m \in M} (y_m - 1),
$$

and we have to maximize  $\mu_{CROSS}(x)$  subject to the constraints (10)-(11) and with the additional crossing count constraints (12).

The reduced cost of a column now contains the term

$$
-\sum_{\mathfrak{m}\in M}b^{\mathfrak{m}}_{r}\lambda_{\mathfrak{m}},
$$

where  $\lambda_m$ ,  $m \in M$  are the dual multipliers of constraints (12).

#### 4.3 The PCON model

IT, MIT and CROSS are all aircraft-based metrics. Another possibility is to use passenger-centric UFs based, for instance, on idle connection time for passenger itineraries with multiple flights.

Let I be the set of all existing passenger connections in the schedule; each of them is defined by a pair of flights  $(f_i, f_j) \in I$ . We define the idle connection time  $\delta_{ij}$  of  $(f_i, f_j) \in I$  as the time between the landing time of flight  $f_1$  and the departure of  $f_2$  minus the minimum passenger connection time (typically 30 minutes), denoted MPC. We assume a constant value for MPC, as assumed by most airlines and in literature (e.g. Lan et al., 2006).

PCON is the UF maximizing the passenger idle time:

$$
\mu_{PCON}=\sum_{(f_i,f_j)\in I}\delta_{ij}.
$$

Given a route r,  $t_r^{\text{dep}}(f_j)$  is the landing time of flight  $f_j$ , which is 0 if  $f_j$  is not covered by route r and the exact departure time of  $f_j$  if route  $r$  covers it. Similarly,  $t_r^{\text{land}}(f_i)$  is the landing time of flight  $f_i$  if it is covered by route r, 0 otherwise. As the covering of all flights is imposed by constraints  $(2)$ , we always have one route  $r$  in the solution with non-zero values of  $t_r^{\text{land}}(f_i)$  or  $t_r^{\text{dep}}(f_j)$ .

In addition to constraints  $(10)-(11)$ , we have to impose non-negativity of each connection time as follows:

$$
\delta_{ij} - (\sum_{r \in \Omega} t_r^{dep}(f_j) - \sum_{r' \in \Omega} t_{r'}^{land}(f_i) - MPC) \le 0 \quad \forall (f_i, f_j) \in I \quad (13)
$$

$$
\delta_{ij} \ge 0 \quad \forall (f_i, f_j) \in I \quad (14)
$$

When maximizing  $\mu_{PCON}$ , the model with constraints (10)-(11) and (13)-(14) ensures that the total passenger connection time is maximized, while satisfying the minimum passenger connection time for all connections I.

From the algorithmic point of view, the structure of the pricing problem is unchanged up to the consideration of additional prices to be collected in the RCESPP algorithm; when taking discretized times for  $t_r^{\text{land}}(f_i)$  and  $t_r^{\text{dep}}(f_j)$ , the collection is similar to the price collection of take-off and landing slots.

#### 4.4 Implementation

The four MRP algorithms, IT, MIT, CROSS and PCON corresponding to the presented UFs, and the recovery algorithm solving the ARP are implemented using the same Column Generation heuristic: column generation is performed only at the root node. The branching scheme is meant to derive an integer solution from the columns obtained at the root node. Furthermore, to speed up computation, we derive three heuristic pricing levels depending on the number of columns found:

- 1. the number of labels to be extended at each node is limited and domination criteria are heuristic, i.e. labels might be erroneously discarded;
- 2. same than level 1, but we increase the number of labels extended at each node;
- 3. the number of labels to extend is unlimited.

When one heuristic level fails to find any column, we proceed to the next level. Eggenberg et al. (forthcoming) show that this leads to a fast heuristic that generates good quality solutions in terms of optimality deviation.

Moreover, when the flight retiming window is smaller than twice its duration for each flight, this procedure leads to the optimal solution of the pricing.

The algorithms are written in C++ using the COIN-OR BCP framework<sup>2</sup>, each algorithm containing around 12,000 lines of code in addition to the COIN-OR BCP framework.

#### 4.5 Simulation Methodology

To validate the above models, we generate different schedules from the same original one with each model using different budget values for C. We then apply a same disruption to each schedule and then run the same recovery algorithm to recover the disrupted schedule.

As some models do not consider passenger connections, it may occur that some of them are no longer feasible after re-timing flights. In such cases, we assume that no ticket using such a connection can be sold, i.e. the passengers are lost and the tickets have to be refunded. The consequence is a loss of revenue, which is the cost of making the schedule more robust/recoverable.

In order to compare the efficiency of different schedules for a same disruption scenario, We adopt a similar approach than Lan et al.

 $^2$ http://www.coin-or.org

(2006): given the original schedule and a disruption characterization, we identify, for each flight, the so called *independent delay* and the propagated delay. The independent part of the delay is single-flight dependent and is, therefore, part of the disruption characterization. The propagated delay is a consequence of the schedule, which is a consequence of the disruption and must be recomputed.

## 5 Computational Results

For the computational results, we use public data provided for the ROADEF Challenge  $2009^3$ . We use the A instance set, i.e. the set of instances used for the Challenge qualification phase.

Each instance is composed of an original schedule and disruption scenario. The original schedule is composed of the existing legs, the routes of each aircraft (including maintenances, that cannot be rescheduled) and the passenger's itineraries. Additionally, there are airport arrival and departure capacities, which are given as upper bounds for each one-hour interval of a typical day. Disruption scenarios are characterized by an operational period prior to the start of the recovery period, for which observed flight delays and flight cancellations are reported. Additionally, mandatory rest periods for aircraft and modified airport capacities at given time slots are also provided.

The recovery algorithm computes new routes for the aircraft and the passengers in order to minimize recovery costs; only flights departing after the start of the recovery period can be rescheduled, all other flights are fixed; the same holds for passenger itineraries. External cost-checker and checker for feasibility are provided, allowing to externally evaluate the solutions according to the real cost-metric.

The qualifying instances A01-A10 are based on the same schedule with 35 airports and 85 planes.

Instances A01-A04 and A06-A09 are single-day schedules with 608 flights and between 36010 and 46619 passenger itineraries, whereas A05 and A10 are a two days schedule with 1216 flights and between 71910 and 95392 passenger itineraries; we refer to them as the 1-day and 2-days instances, respectively.

As discussed in section 4.5, a preprocessing phase is required to apply a disruption scenario to a modified schedule. First of all, for each solution, we remove from the formulation the passengers missing a connection, i.e. with less than 30 minutes connection time, due to

 $^3$ http:// $\verb|challenge.roadef.org/2009/index_en.htm$ 



Model	MIT_20000	CROSS_1000	CROSS_2500	CROSS_5000	CROSS_10000	PCON_1000	PCON_2500	PCON_5000
Used Budget [min]	10025	1000	2500	5000	5980	1000	1250	2500
$#$ Modified Flts	308	109	178	248	255	31.5	26.5	52.5
[min] IT	16750	11880	11415	11450	10965	12815	12960	13670
MIT [min]	3355	690	620	505	460	782.5	807.5	795
<b>CROSS</b>	3410	3494	3517	3530	3519	3447.5	3444	3459.5
PCON [min]	141218	129143	127318	127743	127468	134533	135888	140573
$#$ Lost Pax	438.5	73.5	262.5	366	405.5			
Pax Lost $[\%]$	1.09	0.20	0.67	0.90	1.02	0.00	0.00	0.00
Revenue Loss [%]	3.56	0.71	2.35	3.37	3.65	0.00	0.00	0.00
CPU Time [s]	408	406	412	583	285	757	1058	1073

Table 1: Average a priori statistics on instances A01-A04 and A06-A09.

flight retiming. These lost passengers correspond to the loss of revenue sacrificed to increase the schedule's robustness and recoverability; the number of lost passengers and the corresponding loss of revenue are shown for each instance.

#### 5.1 A priori results

For the presentation of the results, we separate the 1-day instances from the 2-days ones.

The original schedules (as provided in the data set) are labeled  $0r$ ; the schedules obtained by the UFO models are labeled IT, MIT, CROSS and PCON. The UF solutions are followed by a number specifying C in (10), corresponding to total allowed deviation of departure times in minutes. Thus, for example, instance A01 CROSS 1000 corresponds to the solution of instance A01 solved with UF CROSS and a budget  $C = 1000$  minutes.

For each instance, we generate one schedule for five different budgets, namely  $C = 1,000, 2,500, 5,000, 10,000$  and 20,000 minutes respectively; the maximal deviation of a single flight is set to 60 minutes. The complete results are reported in Appendix A.

Table 1 summarizes the average a priori statistics on the 1-day instances and Table 2 for the 2-day instances. Displayed informations are used budget (in minutes), the value of the different UFs for each solution, the statistics of lost passengers (absolute, relative and corresponding relative loss of revenue with respect to the original schedule) and CPU times.

Model	0r	IT_10000	MIT_10000	CROSS 10000	PC0N_1000
Used Budget [min]	$\Omega$	10000	10000	8515	1000
$#$ Modified Flts	0	252	407	424	31
$IT$ [min]	77865	85068	80160	76925	78220
$MIT$ [min]	490	408	1965	140	475
<b>CROSS</b>	6100	6176	6085	6184	6105
PCON [min]	258143	276113	268178	257348	263173
$#$ Lost Pax	0	298	414	671	0
Pax Lost $[\%]$	0.00	0.36	0.50	0.82	0.00
Revenue Loss [%]	0.00	1.30	1.79	2.90	0.00
CPU Time [s]	$<$ 1	10828	5412	6291	41292

Table 2: Average a priori statistics for different models for instances A05 and A10.

First note that Table 1 does not report results for PCON 10000 and for models IT 20000 and CROSS 20000. For model PCON, the algorithm is not able to find a solution different from Or; for the other models, there is no difference between a budget  $C = 10,000$  and  $C = 20,000$ . This is due to the fact that we have a disaggregate bound on retiming for each flight which is independent of C. When C is large enough, the total retiming is limited by the disaggregate bounds before reaching the aggregate bound C, which is the case for models IT 20000 and CROSS 20000.

Table 2 shows that the computational effort for the 2-day instances is increased up to a factor between 13 and 55 with respect to the 1 day instances. The number of aircraft, however, is unchanged, namely 85, and the number of flights is multiplied only by a factor 2. This shows the combinatorial complexity of the problem. Moreover, the UF values are much higher than for the 1-day, explaining why the relative increase of the UFs is lower.

A remarkable point is the number of lost passengers and associated loss of revenue. Indeed, all models except PCON do not consider connections at all. However, for the 1-day instances, the maximal loss of passengers is 1.31% for a single instance and 1.10% in the average. The loss of revenues are slightly higher than the number of lost passengers. The reason is that the misconnected passengers are those with tight connections, which often corresponds to the profile of business passengers, who also pay higher fares. The loss of revenue due to retiming is thus always lower than  $4.3\%$  (3.65% in average) of the original revenue, but note that this is an upper bound: indeed, we do not consider the possibility of attracting additional customers with the connections created in the new schedule.

We also see from Table 1 that the models are able to significantly increase the values of their corresponding UF. We also see that increasing the budget leads to solutions with higher values for the UFs. The increase is not necessarily homogeneous: the value of CROSS is higher for model CROSS 5000 than CROSS 10000, which is due to the fact we are using heuristics.

Interestingly, IT, MIT and PCON are correlated, as solutions with higher values of one of these UFs also have higher value for the others. This is however not always the case, which shows the UFs are not equivalent. Surprisingly, solutions computed with CROSS tend to decrease the value of IT, MIT and PCON but the reverse is not observed.

#### 5.2 Recovery statistics

For instances A05 and A10, there is no operational phase before the start of the recovery period. Therefore, different initial schedules do not affect the disruption scenario. Moreover, for both A05 and A10, the disruption is a severe global capacity reduction: the initial number of departures and arrivals are 3012 and 2892 respectively; in the disrupted scenario, there is a total reduction of 1110 departures and 1051 arrivals, i.e. a total airport capacity reduction of more than 30%. The consequence is a massive flight cancellation, which highly dominates delays and hides differences of the original schedules. The comparison of recovery statistics for these instances is therefore irrelevant and not reported here.

The detailed results after applying the recovery algorithm for the 1-day instances are listed in Appendix B. We report, for each 1-day scenario, the recovery costs as computed by the cost checker provided for the ROADEF Challenge 2009, the total number of canceled flights (including the forced cancellations from the operational period), the number of canceled passengers, which does not include the lost passengers from the scheduling phase (these are removed from the formulation).

The recovery algorithm is exploiting the non-trivial recovery cost structure as expected. The relation between recovery costs and a posteriori statistics such as number of canceled flights, total delay or number of canceled passengers is not uniform. Indeed, these values are not strictly decreasing for decreasing recovery costs.

The reduction of recovery costs is not uniform for a same model with increasing values of budget C. This is not surprising, as the budget allows for better a priori solutions, but does not guarantee the solution

to be appropriate a posteriori for any given scenario. However, some models generate solutions with an impressive recovery cost reduction: model MIT 20000 reduces the recovery costs by 68.5% in average over the 8 instances. In absolute numbers, the highest savings are obtained with model MIT<sub>-20000</sub> for instance A09, saving up to 1.32 Million  $\epsilon$ , which corresponds to a saving of 70.6% compared to the recovery costs for the original schedule. The highest relative saving is 93.0%, again achieved by MIT 20000 for instance A08. CROSS 1000 is the model that has the most often higher recovery costs than  $\texttt{Or},$  namely in 4 out of 8 instances. PCON 2500 is actually the only model higher total recovery costs summed over all scenarios than Or.

CROSS 1000 and Or both have the highest recovery costs for 2 out of 8 instances. In the remaining 4 instances, it is always a different model that has highest recovery costs. The highest increase in recovery costs occurs at instance A07 with model MIT 5000, with an increase of 239,777 $\in$ , i.e. 37.9% more than  $0r$ .

Although we observe significant differences among the different solutions, there is no homogeneous relation between any UF and the recovery statistics: in general, solutions with higher slack have indeed lower recovery costs, but, for example, MIT 2500 has lower recovery costs than MIT 5000.

As the different disruption scenarios are not equally probable, average results are not representative. We therefore analyze the perfor*mance profile* (Dolan and Moré, 2002) of the different models. They represent, for each model s and each instance p, the probability

$$
P(r_{s,p} \leq \tau : 1 \leq s \leq n_s)
$$

of the model's solution to be withing a factor  $\tau$  of the best found solution in the same instance.  $r_{s,p}$  is the value of the solution obtained with model s on instance p divided by the best found solution for instance  $p$  and  $n_s$  is the number of instances solved with model  $s$  (in our case,  $n_s = 8$  for each model).

When  $\tau = 0$ , the value of  $P(r_{s,p} \leq \tau : 1 \leq s \leq n_s)$  is the probability of model s to lead to the best solution. Eventually, when  $\tau$  grows lager, all models s will have a probability  $P(r_{s,p} \leq \tau : 1 \leq s \leq n_s) = 1$ , as all models are able to solve the solution and therefore have a finite value.

Figure 1 shows the performance profile with respect to the recovery costs for Or, IT 10000, MIT 20000, CROSS 5000 and PCON 5000, which correspond to the best solutions for each model. Figure 2 shows more in details the evolution of the performance profiles shown in Figure 1

for a ratio  $\tau \leq 3.5$ 



Figure 1: Performance profile for  $0r$ , IT<sub>-10000</sub>, MIT<sub>-20000</sub>, CR0SS<sub>-5000</sub> and PCON\_5000.



Figure 2: Details for the evolution of the performance curves in Figure 1 for  $\tau \leq 3.5$ .

The best model is clearly MIT 20000, as its probability to be the best model is 0.75. Moreover, it has probability 1 to have recovery costs at most 1.1 times the lowest found solution. Interestingly, for all other models displayed in Figures 1 and 2, there is at least one instance for which the recovery costs are more than 12 times higher than the recovery costs of MIT 20000. We observe also that the second-best model is IT 10000, as is has probability 0.75 to have recovery costs within 1.6 times the lowest found recovery costs. The original solution is the one with lowest probability of being within 3.4 times the best found solution, and also has the highest ratio  $r_{s,p} = 14.27$  for instance A08.

For the models not displayed in Figures 1 and 1, only MIT 10000 is competing with MIT 20000, having probability 0.875 to be within a factor  $\tau = 1.2$  of the best solution; it is also the only solution with ratio  $\tau$  < 10 for instance A08. All other models are below the performance profile of IT\_10000 for  $\tau \leq 2$ . The highest ratio is  $\tau = 14.60$ , obtained with CROSS 1000 for instance A08.

Next, we have to answer the question whether the proposed UFs are significantly correlated or not with the different recovery statistics. Table 3 shows the correlation between the UFs and the different recovery metrics, and Table 4 shows the significance test for the correlations. The statistical test is a bilateral significance test with confidence level  $\alpha = 0.01$  and 166 degrees of liberty (there are 168 observed solutions in total: 8 scenarios, each being evaluated on 21 different solutions). The correlation is significant if the t-value of the test satisfies  $|\text{t min} > 2.606$ .

UF	тт	мтт	CROSS	PCON
Recovery Costs	$-0.371$	$-0.480$	0.052	$-0.269$
<b>Total Delay</b>	$-0.614$	$-0.393$	0.154	$-0.562$
Pax Delay	$-0.550$	$-0.404$	$-0.005$	$-0.269$
Canceled Flights	$-0.004$	$-0.194$	0.152	$-0.026$
Rerouted Pax	$-0.267$	$-0.412$	0.016	$-0.166$
Canceled Pax	$-0.631$	$-0.403$	0.037	$-0.634$

Table 3: Values of the correlation between UF values and recovery statistics.

t-values	тт	мтт	CROSS	PCON
Recovery Costs	$-5.147$	$-7.046$	0.666	$-3.596$
<b>Total Delay</b>	$-10.014$	$-5.510$	2.009	$-8.753$
Pax Delay	$-8.475$	$-5.683$	$-0.067$	$-3.596$
Canceled Flights	$-0.055$	$-2.541$	1.988	$-0.337$
Rerouted Pax	$-3.569$	$-5.822$	0.210	$-2.170$
Canceled Pax	$-10.481$	$-5.669$	0.483	$-10.558$

Table 4: Significance test for the correlation with confidence level  $\alpha = 0.01$ ; the correlation is significant if  $|t| \geq 2.606$ .

Table 3 shows that IT, MIT and PCON have a large negative correlation with all the recovery statistics but the number of canceled flights; CROSS has only low correlation with the metrics. The significance test in Table 4 show that CROSS is not significantly correlated with any of the recovery statistics. Moreover, non of the UFs is significantly correlated with the number of canceled flights.

Interestingly, PCON is not significantly correlated with the number of rerouted passengers. This is somewhat surprising, as the model maximizes the slack for passenger connections and should, therefore, have a higher number of passengers making the connection. A possible explanation is that in  $(13)-(14)$ , we consider the set I of all possible connections. In the data, however, some connections have large connection time (around 6-8 hours) whereas some are tight (30 minutes to 1-2 hours). In the model, however, connection time is considered for both large and tight connections in the same way. An alternative is to restrict I to the set of tight connections, allowing for focusing on the risky connections only. This also simplifies the PCON model, as the number of constraints in  $(13)-(14)$  depends on | I |.

#### 5.3 Synthesis

We solve instances with more than 1200 flights and 85 aircrafts within reasonable computation times. The obtained solutions show that there is a negative correlation between recoverability and IT, MIT and PCON. The correlation is not significant for CROSS, which contradicts the practitioners intuition.

There are two explanations for this. First of all, the results show a reduction of idle time to gain plane crossings, thus also a diminution in the schedule's recoverability. On the other hand, although the recovery algorithm allows for plane swaps, it is the case only for planes of the same fleet. Moreover, CROSS does not differentiate fleets, and assumes homogeneous fleet. To distinguish fleets, we need the meeting point constraints for each fleet type, increasing by another factor the size of the model. This explains why CROSS is not effective in our results. This does, however, not imply that this UF should be discarded, but only that the combination of the CROSS model and our recovery algorithm does not lead to significant increase of recoverability.

The trade-off between loss of revenue at the scheduling phase and savings at the recovery phase is impressive: with MIT 20000, a loss of less than  $143,000\in$  of booking revenue  $(3.57\%)$  enables to save ove  $3.82$  $Mio \in \mathfrak{m}$  terms of recovery costs on the 8 1-day instances.

### 6 Conclusion

In this paper, we present an application of the UFO framework (Eggenberg et al., 2009) to the airline scheduling problem. We present a quantitative simulation to evaluate a solution's performance on real instances,

using an external evaluation tool.

The obtained results show that although our models do not consider any explicit uncertainty characterization, the solutions are able to significantly improve the original solution's recoverability. We prove that an increased idle time improves recoverability of a schedule. In the best case, the total recovery costs over 8 1-day instances can be reduced by more than 3.82 Mio $\epsilon$  which corresponds to a saving of 68.5% with respect to the recovery costs of the original schedule. Additionally, the loss in terms of revenue are small when the models do not consider missed connections: the loss in terms of passenger revenue is always lower than 4.3% of the initial revenue, i.e. less than 22,100 $\in$ ; however, these losses do not consider the possibily of additional bookings on the new connections created in the schedule.

This study opens different research directions. From the computational part, the developed algorithms have still potential for improvements: replace the heuristic by the exact version of the algorithm, improve convergence speed with smart branching decisions, etc. The recovery algorithm would also benefit from an efficient generator of repositioning flights.

In terms of application, other UFs and the combination with different recovery algorithms should be tested in order to better understand the relations between UFs and recoverability; the relation between UFs and different recovery algorithms; the correlation between the different UFs; the efficiency of UFs for different airlines. Finally, the simulations should be extended considering crews and crew recovery, as this is a crucial part in airline operations; this would allow to test crew-based UFs.

### References

- Ageeva, Y. (2000). Approaches to incorporating robustness into airline scheduling, Master's thesis, Massachusetts Institute of Technology.
- Argüello, M., Bard, J. and Yu, G. (1997). A grasp for aircraft routing in response to groundings and delays, Journal of Combinatorial Optimization 5: 211–228.
- Bard, J., Yu, G. and Argüello, F. (2001). Optimizing aircraft routings in response to groundings and delays, IIE Transactions 33: 931– 947.
- Barnhart, C., Johnson, E., Nemhauser, G., Savelsbergh, M. and Vance, P. (1998a). Branch-and-price: Column generation for solving huge integer programs, Operations Research 46(3): 316–329.
- Bian, F., Burke, E., Jain, S., Kendall, G., Koole, G., Landa Silvaand, J., Mulder, J., Paelinck, M., Reeves, C., Rusdi, I. and Suleman, M. (2005). Measuring the robustness of airline fleet schedules, Multidisciplinary Scheduling: Theory and Applications, Springer US, pp. 381–392.
- Clausen, J., Larsen, A. and Larsen, J. (forthcoming). Disruption management in the airline industry - concepts, models and methods, Computers & Operations Research doi:10.1016/j.cor.2009.03.027.

URL: http://www.agifors.org/award/submissions2008/Weide paper.pdf

- Dolan, E. and Moré, J. (2002). Benchmarking optimization software with performance profiles, Mathematical Programming Ser. A 91: 201–213.
- Eggenberg, N., Salani, M. and Bierlaire, M. (2009). Uncertainty feature optimization: an implicit paradigm for problems with noisy data, Technical Report TRANSP-OR 080829, Ecole Polytechnique Fédérale de Lausanne, Switzerland.
- Eggenberg, N., Salani, M. and Bierlaire, M. (forthcoming). Constraint-specific recovery networks for solving airline recovery problems, Computers & Operations Research doi:10.1016/j.cor.2009.08.006.
- Kohl, N., Larsen, A., Larsen, J., Ross, A. and Tiourine, S. (2007). Airline disruption management - perspectives, experiences and outlook, Journal of Air Transport Management 13(3): 149–162.
- Lan, S., Clarke, J.-P. and Barnhart, C. (2006). Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions, Transportation Science 40: 15–28.
- Righini, G. and Salani, M. (2006). Symmetry helps: Bounded bidirectional dynamic programming for the elementary shortest path problem with resource constraints, Discrete Optimization  $3(3): 255 - 273.$
- Rosenberger, J., Johnson, E. and Nemhauser, G. (2004). A robust fleet assignment model with hub isolation and short cycles, Transportation Science 38(3): 357–368.
- Rosenberger, J., Schaefer, A., Golldsman, D., Johnson, E., Kleywegt, A. and Nemhauser, G. (2003a). A stochastic model of airline operations, Transportation Science 36(4): 357–377.
- Shebalov, S. and Klabjan, D. (2006). Robust airline scheduling: Moveup crews, Tranportation Science 40(3): 300–312.
- Thengvall, B. G., Bard, J. F. and Yu, G. (2000). Balancing user preferences for aircraft schedule recovery during irregular operations, IIE Transactions 32(3): 181–193.
- Weide, O. (2009). Robust and Integrated Airline Scheduling, PhD thesis, The University of Auckland, New Zealand.
- Yen, J. W. and Birge, J. R. (2006). A stochastic programming approach to the airline crew scheduling problem, Transportation Science 40: 3–14.

# A Complete proactive statistics

Tables 5-12 report the a priori statistics for the 1-day instances (A01- A04 and A06-A09), and Tables 13 and 14 for the 2-day instances A05 and A10.

# B Complete recovery statistics

Tables 15-22 report the recovery statistics for the 1-day instances (A01- A04 and A06-A09).

















































# Rerouted Psg | 597 | 571 | 508 | 492 | 473 | 541 | 572 | 572 | 572 | 572























