Pre-Tactical Planning of Runway Utilization Under Uncertainty: Optimization and Validation

Norbert Fürstenau†, Andreas Heidt*, Manu Kapolke*, Frauke Liers*, Monika Mittendorf† and Christian Weiß†

*Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Economics, Discrete Optimization, Mathematics
Cauerstr. 11, 91058 Erlangen, Germany
Email: name.surname@fau.de

†German Aerospace Center (DLR)
Institute of Flight Guidance
Lilienthalplatz 7, 38108 Braunschweig, Germany
Email: name.surname@dlr.de

***Abstract*—Efficient planning of runway utilization is one of the main challenges in Air Traffic Management (ATM). In a previous paper, we developed a specific optimization approach for the pre-tactical planning phase that reduces complexity by omitting unnecessary information. Further, we investigated the impact of disturbances on our solutions, since in reality uncertainty and inaccuracy almost always lead to deviations from actual plans. In this paper, we now present approaches to incorporate uncertainty directly in our model in order to achieve a stabilization with respect to changes in the data. Namely, we use techniques from robust optimization and stochastic optimization. Further, we analyze real-world data from a large German airport to obtain realistic delay distributions, which turn out to be two-parametric Γ-distributions. We describe a simulation environment and test our new solution methods against standard algorithms (e.g., First-Come-First-Serve). The encouraging results show that our approaches significantly reduce the number of necessary replannings.**

I. Introduction

ATM systems are driven by economic interests of the participating stakeholders and, therefore, are performance oriented. As possibilities of enlarging airport capacities are limited, one has to enhance the utilization of existing capacities to meet the continuous growth of traffic demand. The runway system is the main element that combines airside and groundside of the ATM System. Therefore, it is crucial for the performance of the whole ATM System that the traffic on a runway is planned efficiently. Such planning is one of the main challenges in ATM. Uncertainty, inaccuracy and non-determinism almost always lead to deviations from the actual plan or schedule. A typical strategy to deal with these changes is a regular re-computation or update of the schedule. These adjustments are performed in hindsight, i.e. after the actual change in the data occurred. The challenge is to incorporate uncertainty into the initial computation of the plans so that these plans are robust with respect to changes in the data, leading to a better utilization of resources, more stable plans and a more efficient support for ATM controllers and stakeholders. Incorporating uncertainty into the ATM planning procedures further makes the total ATM System more resilient, because the impact of disturbances and the propagation of this impact through the system is reduced.

In the present paper, we investigate the problem of optimizing runway utilization under uncertainty. We incorporate uncertainties into the initial plan in order to retain its feasibility despite changes in the data. We focus on the pre-tactical planning phase, i.e. we assume the actual planning time to be several hours, or at least 30 minutes, prior to actual arrival/departure times. In our contribution to the SESAR Innovation Days 2014 [6], we developed an appropriate mixed integer program (MIP) for this particular planning phase. The basic idea was that in pre-tactical planning we can reduce the complexity of the problem by not determining exact arrival times for each aircraft, but assign aircraft to time windows of a given size. Afterwards, the impact of disturbances on the deterministic solutions was investigated. The results showed that it is crucial to enrich the optimization approach by protection against uncertainties, in order to produce less necessary replanning. In the current paper, we thus incorporate uncertainties directly into the model by using techniques from robust and stochastic optimization. These techniques are then tested within a simulation environment.

The remainder of this paper is organized as follows: In Section II, we briefly describe the pre-tactical runway optimization model which we developed in [6]. Afterwards, we present our robust and stochastic optimization approaches to incorporate uncertainties into this model in Section III. To test our solution methods in a more realistic setting, we analyze real-world delay data from a large German airport in Section IV (extending the descriptions in [6]), where we also describe our simulation environment and discuss the obtained validation results for our new optimization approaches, tested against standard algorithms. Finally, we conclude in Section V.

II. The Modeling

For a detailed description of our nominal optimization model, see [6]. We model the pre-tactical planning phase by

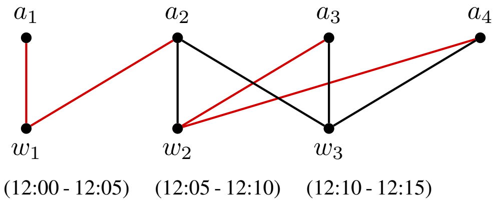

Fig. 1. Assignment graph. Red edges show a possible assignment: aircraft a_1 and a_2 are assigned to time window w_1, a_3 and a_4 are assigned to w_2.

assigning time windows to aircraft. We consider single-mode runways with only arriving aircraft. In our modeling approach we claim that each aircraft has to receive exactly one time window as each aircraft has to be scheduled. On the other hand, the number of aircraft that can be assigned to one time window depends on its size and the weight classes of the aircraft. The underlying idea is that, contrary to tactical planning, we don't need to determine arrival times to the minute yet, because we are several hours (or at least 30 minutes) prior to the first scheduled time. Thus, the exact arrival sequences within the time windows can be decided later.

For each aircraft, we consider several corresponding times:

- *Scheduled time of arrival* (ST): a fix time that yields a benchmark to identify delay and earliness of the aircraft. This may be the time the passenger finds on his flight ticket.
- *Earliest time of arrival* (ET): depends on operational conditions (and on the impact of disturbances).
- *Latest time of arrival* (LT): latest time the aircraft can land without holdings. It depends on the earliest time ET and on the actual planning time (or start time, respectively, if the aircraft is still on the ground).
- *Maximal latest time of arrival* (maxLT): a hard condition for landing which is calculated with respect to physical, operational and other relevant conditions (for instance, amount of fuel, prioritization, etc.).

Those times further determine the corresponding time windows ST_W, ET_W, LT_W and maxLT_W for each aircraft. Each aircraft can be assigned to all time windows between ET_W and maxLT_W. To model the problem mathematically, we consider a bipartite assignment graph $G = (A \cup W, E)$ consisting of a vertex set A of aircraft and a vertex set W of time windows of a given size in a given time period (ordered chronologically). An edge $(i, j) \in E$ corresponds to a possible assignment of aircraft i to time window j. In Figure 1 we see a small example of a bipartite graph with a possible assignment of aircraft $a_1, \dots, a_4 \in A$ to time windows $w_1, w_2, w_3 \in W$.

Our objective is the minimization of delay and earliness, respectively. Delay is penalized quadratically for reasons of fairness (e.g., a solution in which one aircraft has a delay of six time windows is worse than a solution in which two aircraft have a delay of three time windows each). Earliness is penalized linearly. If the assigned time window is after the LT_W (i.e. between LT and maxLT), we add an extra penalization term.

TABLE I
MINIMUM SEPARATION TIMES (IN SECONDS)

Predecessor \ Successor	Heavy	Medium	Light
Heavy	100	125	150
Medium	75	75	125
Light	75	75	75

The constraints in our MIP consist of general assignment constraints and the modeling of minimal time distance requirement. Those minimum separation times between two consecutive aircraft depend on their corresponding weight classes. Hereof, we consider three different aircraft categories (*Light*, *Medium* and *Heavy*) and use Table I ([9]).

III. Incorporating Uncertainties

In this section, we want to incorporate uncertainty into the model to receive a robustification of our solution plan. In general, robustification means to ensure that deviations in the input data do not have a large impact on the solution. Considering the optimal solution of the nominal problem, i.e. the problem where uncertainties are ignored, small deviations in the input data could have the effect that the nominal optimum becomes infeasible for the disturbed problem, i.e. the problem where the input data suffers from deviations. Computational results that showed a significant impact of disturbances on our nominal solutions can be found in [6].

In mathematics, there exist different approaches to handle uncertainty in optimization. In *stochastic optimization* (e.g. [10]) the goal is to describe the uncertainty by probability distributions. Knowing these distributions, one can then optimize the expected values. A second approach to the problem of modelling uncertainty is located in *robust optimization* (e.g. [2], [3]), where the goal is to immunize against predefined worst-case scenarios. In contrast to stochastic optimization, the probability distributions of the uncertainties do not need to be known. However, one has to predefine uncertainty sets that determine the values of the uncertain parameters against which the optimization problem has to be protected. The task is to find robust feasible solutions, i.e. solutions that are feasible for all parameter values in the uncertainty set. Among all robust feasible solutions, the robust optimal solutions are those with the best guaranteed objective function values.

1) Robust Optimization Approach: In the setting for our model described in section II (and precisely in [6]), the uncertain parameters are the ET windows ET_W and, dependent on those, LT_W and maxLT_W. Hence, we have to predefine an uncertainty set for each aircraft. Therefore, we have to chose deviations of the earliest time we want to be protected against. For each aircraft this yields an interval of possible earliest times and thus a set of possible ET_W's. These ET_W's also determine the possible LT_W's.

Now, we actually solve our optimization model from section II. But in the robust approach we assume an assignment graph that only contains edges corresponding to assignments which are feasible for every realization of our chosen uncertainty

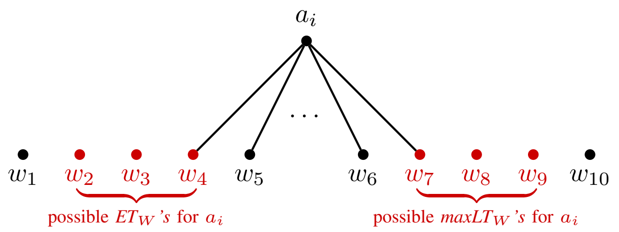

Fig. 2. Possible assignments for an aircraft a_i in the robust model

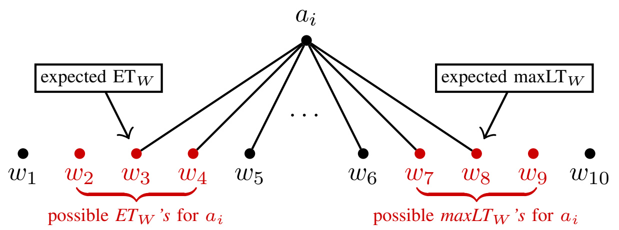

Fig. 3. Possible assignments for an aircraft a_i in the stochastic model

set. An example of feasible assignments for an aircraft in the robust model is illustrated in Figure 2. As mentioned, the robust model assumes the worst-case, i.e. the extreme cases for earliest time (w_4) and maximal latest time (w_7) in the predefined uncertainty set are taken into account, whereas the other time windows which lay within the uncertainty set for both times (w_2, w_3, w_8, w_9) are forbidden.

2) Stochastic Optimization Approach: We follow a single-stage stochastic optimization approach in which we optimize over all assignments which are "expected to be possible" dependent on the underlying probability distribution. Therefore, we consider the expected values for ET and maxLT for each aircraft, or the corresponding time windows, respectively. Afterwards, we optimize the obtained "expected scenario", i.e. we solve our mathematical model described above with edges in the assignment graph that correspond to the feasible assignments in this scenario. In Figure 3 we show an example of feasible assignments in the expected scenario for one aircraft.

A well-known combination of robust and stochastic methods is to determine the uncertainty set in the robust approach using stochastic values. The idea is that the uncertain parameter does not deviate from its expected value by more than k times of its standard deviation. This can help to decide which boundaries should be chosen for the time window uncertainty sets. Note that $k = 0$ yields the described stochastic approach.

So far, in this paper we have described a mathematical approach for optimizing runway utilization in the pre-tactical planning phase. Further we have enhanced our developed optimization model by incorporating uncertainties in different ways (robust and stochastic). In the following section, we now analyze real-world disturbances from our database from a large German airport. Afterwards, we describe a simulation environment and test our developed approaches against standard algorithm with those realistic disturbances.

IV. Validation Experiments

Understanding and modeling the statistics, dynamics, and propagation of air-traffic arrival and departure delays is a prerequisite of any attempt to optimize the punctuality of schedules and airport capacity, and minimizing necessary buffer times for required robustness of performance (e.g. [12], [13]). That is why for validating the new scheduling models by means of Monte Carlo simulations we start with the design of an appropriate stochastic delay model.

A. Stochastic Delay Model

For initial validation of the new stochastic and robust optimization algorithms for aircraft sequencing in the pre-tactical phase we investigate a simple stochastic arrival and departure delay model that is tested by means of empirical delay data from a large German airport. Recently, Caccavale et al. [4] presented a model for simulating inbound traffic over a congested hub termed "pre-scheduled random arrivals" (PSRA) where they defined the actual arrival time $t_i^{\mathrm{ATA}} := t_i$ by a close to Poisson process with mean inter-arrival times $1/\lambda$ of clients in a queueing line:

$$t_i = \frac{i}{\lambda} + \epsilon_i, \quad i = 1, \ldots, n \in \mathbb{Z} \tag{1}$$

The model is represented by a continuous probability density function (PDF) $f_\epsilon(t)$ of the random arrival time variable ϵ with finite standard deviation σ and zero mean, without loss of generality. $1/\lambda$ = expected inter-arrival time between two consecutive aircraft, $1/\lambda = \mu = \langle \Delta t_i^{\mathrm{ATA}} \rangle = \langle t_i^{\mathrm{ATA}} - t_{i-1}^{\mathrm{ATA}} \rangle$, with actual arrival times t_i^{ATA} = actual in-block time AIBT, in what follows. Guadagni et al. [7] prove that this process converges to the memoryless one-parametric Poisson process for large σ. This approach overcomes the often used assumption of uncorrelated arrivals as precondition of the Poisson process, i.e. exponentially distributed inter-arrival times $\Delta t_i^{\mathrm{ATA}}$. Empirical histograms of delay data exhibit a pronounced non-symmetry (e.g. [12]) that was modeled by Wu [14] by means of the two-parametric Beta-probability density function (limited to the open $(0, 1)$ interval). For our purpose the family of two-parametric Gamma (Γ)-PDF's (limited to $\mathbb{R}^+$, with shape and scaling parameters a, b) appears more appropriate as analytical model, because it contains the exponential distribution of the Poisson process as a special case ([5]).

A realistic model of arrival delays, in addition to the asymmetry has to include a significant amount of early arrivals, i.e. delay $t^D < 0$. Furthermore, besides the statistics of the sequence of all different arrivals a_i (different flights) during single days of operation (single day statistics) also single flight (=airline) statistics (e.g. all arrivals j of the same flight a_{ij} over a time interval of e.g. half a year) have to be modeled ([1]). The delay statistics naturally exhibits daily, weekly, and seasonal periodicities and trends, i.e. nonstationary behavior. Consequently any realistic model has to be a combination of deterministic and random components ([1], [12]) which is one reason for the inappropriateness of the Poisson model that represents maximum randomness. For taking into account early arrivals ($t^D < 0$) each histogram data set has to be transferred into $\mathbb{R}^+$ by subtracting the minimum delay (minimum earliness $t^D_{\min}$) before data fitting with the Γ-model. The Γ-PDF as a generalization of the Poisson process of inter-arrival times (t) may be parametrized by the shape parameter a and the mean τ.

$$f(t;\tau,a) = \left(\frac{a}{\tau}\right)^a \frac{t^{a-1}}{\Gamma(a)}\, e^{-\frac{at}{\tau}} \tag{2}$$

with normalized time scale t/τ, scaling parameter b defined via $\tau = a \cdot b$, and the 2nd and 3rd (central) moments $\mu_2 = \text{variance} = \sigma^2 = \frac{\tau^2}{a} = a \cdot b^2$, $\mu_3 = 2ab^3 = 2\sigma^2 b = \frac{2\tau^3}{a^2}$, with skewness $\gamma = \frac{\mu_3}{\sigma^3} = \frac{2}{\sqrt{a}}$, and coefficient of variation $cv = \frac{\sigma}{\tau} = \frac{1}{\sqrt{a}}$ independent of τ. A residual linear correlation $cv \sim \tau$ of empirical PDF's would result in an inverse power law $a(b)$-anticorrelation. For $a = 1$, (2) reduces to the Poisson process of maximum randomness, i.e. exponential t-distribution. For $a < 1$, (2) models a process with larger variance than the random process due to clustering, i.e. non-independent clustered events. For large $a > 1$, with the Γ-PDF approaches a (τ, σ)-Normal distribution.

The Γ-model may be related to the PSRA model by splitting the average inter-arrival time $\mu = \frac{1}{\lambda}$ into the deterministic (schedule) part μ^{STA} and the stochastic delay contribution $\mu = \mu^D + \mu^{\text{STA}}$, $\mu^D = \tau + t^D_{\min}$ (usually $t^D_{\min} < 0$):

$$t^D_i = \mu^D + \epsilon_i = \tau + t^D_{\min} + \epsilon_i \tag{3}$$

where ϵ_i collects the random contributions from μ_2 and μ_3. The analysis of empirical arrival and departure delay histograms in the following section IV-B together with Monte Carlo (MC) computer experiments with the different optimizer models in section IV-C in fact indicate Γ-models to provide reasonable approximations for the arrival and departure delay statistics as one usable metric for the optimizer performance differences, with characteristic deviations from Γ-PDF due to the optimization (see Figure 7).

B. Analysis of Empirical Arrival- and Departure Delays and Derivation of Disturbance Statistics

In this section we model the empirical arrival and departure delays of flights a_{ij} ($i = 1, \ldots, m$) with a stochastic Γ-process according to (2) and (3), with delays = random deviations from scheduled arrival times (STA, flight plan), and we derive an empirical disturbance statistics for use with the MC-computer experiments. As proposed by Abdel-Aty et al. [1] we analyze and model daily delays observed within the time series of all flights a_{ij} ($i = 1, \ldots, m > 200$) during full days of operation, as well as delay data from a selection of single flights a_j over a couple of months (with $j = 1, \ldots, n \geq 150$ monitored arrivals or departures).

Figure 4 shows an example of arrival delay probabilities $f(\text{ATA(=AIBT)} - \text{STA})$ for a single full day. We also analysed

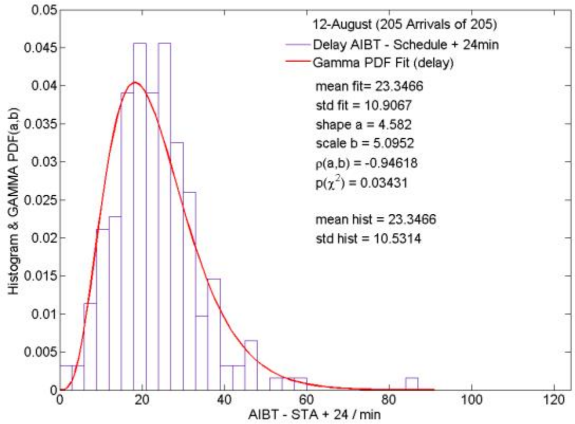

Fig. 4. Example of empirical arrival-delay histogram (AIBT−STA−$t^D_{\min}$ (earliness: $t^D_{\min} = -24$ min)) from the data base at a large German airport (shifted into $\mathbb{R}^+$) with Γ-PDF fits yielding a, b estimates. Full-day (17 hours) traffic with 205 evaluated arrivals.

a sample of 33 flights (different callsigns) with ≥ 150 arrivals each per half year (out of 1384 within 7 - 12/2013). The χ^2-tests of the maximum likelihood (ML) Γ-fits to the empirical delay histograms differ significantly between single days as well as between single flights. This is no surprise, of course, due to the neglection of any deterministic effect (correlations between flight arrival times or delays depending on traffic density, previous flight delay, etc.).

The figure legend provides the fit results for the parameter estimates a, b with Γ-mean τ (same value for empirical histogram and ML-estimate), a-b correlation coefficient, and χ^2-test of Γ-hypothesis (0-hypothesis rejection for $p < 5\%$). The fit example in this case in fact formally should be rejected at the $p = 5\%$ level, basically due to the deviations around zero delay (AIBT−STA−$t^D_{\min} = 24$ min). Besides the neglection of the above mentioned deterministic effects, this deviation around $t^D = 0$ can be explained by active ATC interventions to minimize delays (replaced by the algorithmic scheduling optimization in the following section IV-C). Nevertheless we obtained many examples without 0-hypothesis rejection, i.e. $p(\chi^2) > 5\%$. The average fit parameter estimates for the 33 single flights a_i are (± 1 stddev): $\langle a \rangle = 3.5(1.3)$, $\langle b \rangle = 8.7(3.4)$, $\langle \tau \rangle = 27.5(7)$ min, with average minimum earliness $\langle t^D_{\min} \rangle = -23.9(8.8)$ min (transformation into $\mathbb{R}^+$ by $t^D_{\min}(a_i)$ for each single fit), yielding an average arrival delay of $\langle \mu^D \rangle^{\text{Arr}} := \langle \tau \rangle - \langle t^D_{\min} \rangle \approx 3.6(11)$ min, with stderror of mean $\epsilon = 2$ min.

For the simulations in section IV-C we will use departure delays (ATD−STD) as the only disturbance during the flight. This is motivated by the fact that according to Eurocontrol statistics (see Performance Review Report [11]) departure delays represent the main source of arrival delays. Figure 5 depicts an example departure delay histogram with Γ-fit to empirical data of one from 46 single flights with ≥ 150 departures. They were obtained out of 1579 analysable flights

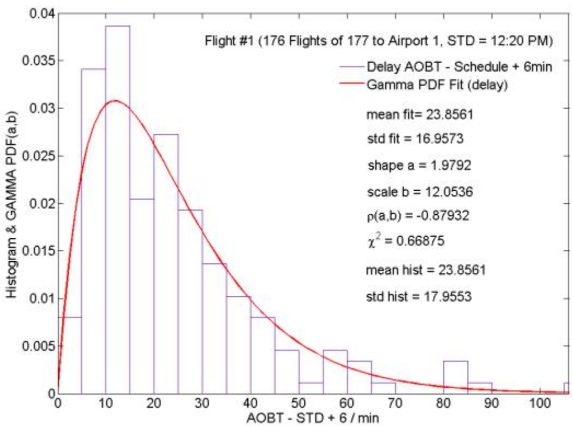

Fig. 5. Example of empirical departure delay histogram (time shifted to $\mathbb{R}^+$ by -6 min earliness) with maximum likelihood Γ-PDF fit for all 176 departures of a single flight number over 6 months. χ^2-test is significant at 5% level. Parameter estimates of fit in the legend.

a_i from altogether 32604 departures during a 6-months time span.

The corresponding average departure delay parameters of $f(\text{ATD(=AOBT)} - \text{STD})$ for all 46 flights are $\langle a \rangle = 2.5(0.8)$, $\langle b \rangle = 8(3.4)$, $\langle \tau \rangle = 18.2(5.4)$ min, $\langle t^D_{\text{min}} \rangle = -10.9(4.1)$ min, yielding an average departure delay $\langle \mu^D \rangle^{\text{Dpt}} := \langle \tau \rangle + \langle t^D_{\text{min}} \rangle \approx 7.3(6.6)$ min. Comparing this value with the average of the 33 mean arrival delays yields the departure delays about 4 min larger. This difference compares well with statistics reported in [11]. Also the larger variation of the mean arrival delays $\sigma(\mu^D)^{\text{Arr}} \approx 7.0$ min as compared to the mean departure delay variation $\sigma(\mu^D)^{\text{Dpt}} \approx 5$ min compares well with PRR-results, although this is partly explained by the different sample size 33/46. Because no sufficient empirical data from departure delays from the origin airports of the flight were available we use the departure delays of the destination airport as representative departure disturbance value for the MC-simulations with the different scheduling optimization algorithms and models.

Derived from an empirical data set as used for Figure 4, the tuple (take-off time TOT, ET, STA, latest and absolute latest times LT, LTmax)$_i$ from a well-defined series of 209 flights of a full single-day of traffic (17 h time span) was used as input for the MC-simulations of the standard traffic scenario (S1). Because the corresponding average traffic density of ca. 12 flights/h was low compared to the published capacity of 27 arrivals/h (plus 27 departures/h) we created in addition a dense scenario (S2) for testing the optimizers. The whole traffic of 209 A/C of the empirical standard scenario in this case is compressed to a reduced time span (8 h from originally 17 h, starting at 6:00) yielding a traffic density of 26 arrivals / h near the capacity limit. This was realized in such a way that all flights with arrival times $< t_0 + 8$ h remain unchanged and the rest up $t_0 + 17$ h is put in between these flights with correspondingly shifted (ET, STA, LT, LTmax)$-$times.

Furthermore each flight a_i is characterized by its individual weight class that determines its minimum separation distance from the previous flight a_{i-1} according to Table I (section II). Because the original scenario contained only 8 H-class A/C we increased the number (and traffic complexity) to 24 by changing those M-class with long flight distance ($>$ 1500 km) into H-class. The modified empirical scenario (S6.2, S7.2) contained 24 (11.5%) H-class A/C, 14 (6.7%) L-class, and 171 (81.8%) M-class A/C.

C. Monte Carlo Simulations

1) General Aspects: For calculating and updating the individual target times TT for each flight a_i of the full day schedule (with ET $\leq$ TT $\leq$ LT $<$ LTmax), the computer experiments used a simplified time-based trajectory model defined by the individual earliest and latest times of arrival (ET, LT, LTmax). For the pre-tactical phase before departure ET = constant, LT = LTmax. After the departure ET converges to TT with increments $\sim \Delta t^{\text{Sim}}(\text{TT}-\text{ET})/(\text{TT}-t^{\text{Sim}})$, and the interval (LT$-$ET) as function of simulation time t^{Sim} decreases linearly according to $(\text{ET} - t^{\text{Sim}})/2$, with some modifications during final approach $<$ 30 min before arrival which however, are not of interest within the present work (TT$-t^{\text{Sim}} >$ 30 min, (TT$-$ET) $<$ ca. 5 min).

Target Time TT for each simulation time step Δt^{Sim} (= 4 min) is the optimization result with regard to minimizing for the whole daily arrival sequence the deviations from the individual schedules STA(a_i), or alternatively from ET(a_i) (see Discussion section IV-D), based on the specific objective or cost function (see above). An update of optimized a_i-sequences is calculated for each Δt^{Sim}, and the daily sequence will undergo changes as long as new flights are starting from their respective departure airports, with the individual departure delay drawn from the same average Γ-PDF (a = 2.5, b = 8, τ = 18.2 min; see previous section) and shifted back to the delay scale μ^D. Typically, for 17 hours of daily operation of our empirical dataset we have ca. 260 simulation steps per MC-run. Runtime depends on the traffic density, time of operation and sequencing algorithm (optimizer), and varies between (typically) 1 s (first-come first-serve rule (FCFS) = no optimization), 15 s for the three MIP models (discrete assignment windows = 10 min), and 200 s for Take Select 8-2. With 200 MC-runs per experiment we typically have up to several hours of simulation time, depending on the specific optimizer model and scenario. The simulations run on a high performance PC with 2xIntel 64 Bit E5645 12 core processors (24 cores with hyperthreading "on"), 2.4 GHz, 24 GB RAM.

2) Baseline Simulations: In order to establish a baseline, the MC-simulations as a first step were performed without considering a-priori knowledge of disturbance. The three corresponding baseline simulations used the First-Come-First-Serve rule (FCFS), a standard optimizer (Take Select 8-2 ([8]), requiring a monotonous version of the objective function with zero cost for early arrivals), and the nominal model (developed in [6]) based on the same Mixed Integer (Gurobi) discrete optimizer that was also employed with the new stochastic and robust models.

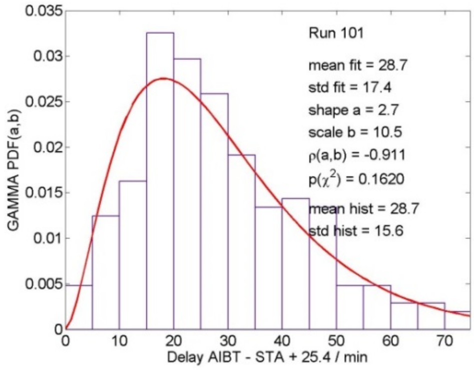

Fig. 6. Example histogram and Γ-PDF fit of baseline MC-simulations with FCFS method (time shifted to $\mathbb{R}^+$ by $t^D_{\min} = 25.4$ min). Single MC-run = single day of operation (8 hours, 209 arrivals) depicting statistics of the dense arrival sequence (scenario S7.2). $p > 5\%$ and low covariance anticorrelation $\rho(a, b)$ suggest acceptance of Γ-fit.

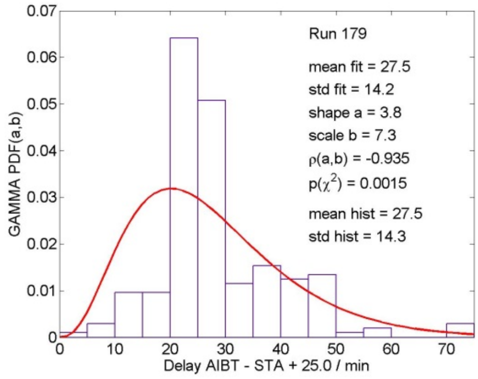

Fig. 7. Example single MC087-run histogram (time shifted to $\mathbb{R}^+$ by $t^D_{\min} = 25$ min) and Γ-PDF fit of baseline MC-simulations with scheduling using the nominal model with additional intra-window sequencing, depicting statistics of the dense arrival sequence (Scenario S7.2). Γ-PDF fit not significant $p < 5\%$.

Figure 6 depicts an MC-simulation (MC057: S7.2) with the FCFS rule (i.e. no optimization) as an example for a single day (= single run) delay statistics for all 209 flights of 8 hrs of operations. The figure shows the delay histogram with Γ-PDF fit that may be compared with the empirical PDF of Figure 4. For most runs the Γ-PDF fits to the delay histograms exhibit good χ^2-test results (no rejection of 0-hypothesis at 95% confidence level).

A corresponding result is obtained for the single flights a_i-analysis with 200 repeated arrivals each. Table II summarizes the baseline results of the continuous time MC-simulations. The ⟨average⟩ parameters of the 209 individual histograms (μ, σ) with Γ-PDF fits for each single flight exhibit results similar to the single days case. The latter numbers (averages $\langle\cdot\rangle$ of fit-parameters (a, b; τ, $\sigma = \tau/\sqrt{a}$), with mean standard deviations $(\cdot)$, times in minutes) for the 200 MC-runs in each case are contained in Table II.

As expected, the results of FCFS and TS8-2 already show that with higher traffic load the use of an optimization algorithm becomes more advantageous both with regard to the number of re-schedulings and absolute mean delay. The general agreement on average of mean delays (μ from Γ-PDF fit), as obtained from single day and single flight delays proves the consistency of the analysis of the $200 \cdot 209 \approx 40000$ entries MC-data tables although, for the TS8-2 optimizer, the single flight analysis (in contrast to the inter-run variation of single day evaluation) exhibits significant inter-individual scattering. We also observe a tendency towards more symmetric PDF's (larger shape parameter a, and delay $\tau = ab$, smaller skewness $2/\sqrt{a}$) with increasing traffic density.

Because preliminary tests with the discrete (Gurobi) MIP-optimizer ([6]) as well as initial MC-simulations showed mainly the higher traffic load (scenario S7.2) to provide sufficient computational demand for evaluating performance differences with the different models, we did put the focus on this condition. The nominal model using the MIP (Gurobi) optimizer provides a third baseline for comparison with the stochastic and robust model results below. According to the preliminary tests we selected a $w = 10$ min window for discretizing the full time span (ca. 8 h) of the dense scenario S7.2. Clearly this discretization does not provide sufficient time resolution for generating a delay PDF and testing the Γ-model. Also the average delays can only provide a value biased to earliness because the early-edges of the windows represented the arrival times for flights (a_i) assigned to the respective windows.

For the purpose of comparing the delay distribution with the two previous continuous baseline scheduling approaches we depict in Figure 7 one MC-run (out of 200 from MC087) where we use the intra-window scheduling (according to separation matrix) for calculating a quasi continuous sequence.

Although the Γ-PDF fit is not significant, the nominal time-window based scheduling achieves results with additional intra-window separation which is comparable to the empirical data. In many other of the 200 runs the histograms reflect the discrete 10 min window sequencing by empty 10 min intervals within the distribution. The strong deviation between 20 - 30 min ($t^D - t^D_{\min} = 25$ min corresponds to delay = 0) is observed for all runs. It reflects the action of the optimizer trying to minimize delays around $t^D = \text{AIBT} - \text{STA} = 0$.

3) Stochastic and Robust Optimization: The results obtained with the nominal model may be directly compared with those of the new models which optimize the arrival sequences by taking into account statistical a-priori knowledge. This was expressed as a shift of the ET, LT, LTmax values as derived from the known delay statistics, represented by the empirical first two central moments ($\mu = \tau + t^D_{\min}$, $\sigma = \tau/\sqrt{a}$): ET:= ET $+\mu + k\sigma$, LTmax:= LTmax $+\mu - k\sigma$, with $k = 0$ for single step stochastic, $k = 1$ for the robust model, and $(\mu, \sigma) := (\mu^D, \sigma^D) = (7.3, 11.9)$ minutes (see above). In Table III we compare for the dense traffic scenario S7.2 and window $w = 10$ min the averages of shape parameters, delays $\langle\mu(\text{std})\rangle$ and number of re-schedulings $\langle rs\#(\text{std})\rangle$ for the single step stochastic (MC088) and robust models (MC089)

TABLE II
CONTINUOUS TIME BASELINE SIMULATIONS WITH PARAMETER AVERAGES OVER 200 RUNS (=SINGLE DAYS) EACH

Model (Scenario)	$\langle RT(\text{std})\rangle/s$	$\langle a\rangle$	$\langle\tau(\text{std})\rangle/\text{min}$	$\langle t^D_{\min}\rangle$	$\langle\mu^D\rangle$	$\langle rs\#(\text{std})\rangle$
FCFS (6.2)	2.2 (1.8)	1.7 (0.2)	23 (1)	-30 (0.5)	-7.0 (1)	2.2 (0.9)
FCFS (7.2)	1.1 (0.4)	2.5 (0.4)	34 (20)	-30 (2)	5.3 (5)	12.7 (5.6)
TS8-2 (6.2)	40 (2)	2.4 (0.3)	24 (2)	-29 (1)	-5.0 (5)	2.5 (1.1)
TS8-2 (7.2)	203 (63)	3.5 (0.6)	29 (4)	-25 (2)	3.5 (6)	8.1 (3.2)

and the three baseline models.

The number of re-schedulings of all three MIP-models decreased significantly relative to the TS8-2 optimizer while FCFS (no optimization) exhibits the highest value. In fact, the robust model protection of optimized TT against departure time disturbance through shifted boundaries of $\pm 1\sigma$ for ET/LTmax, respectively, stabilizes the sequencing significantly more than nominal and stochastic. Namely, we have only half as many re-schedulings using the robust model compared with the nominal one. Hence, we achieve a substantial stabilization of our plans, which is exactly the promise of robust optimization. However, obviously even the robust model produces (very small) $rs\#$-values, which is due to the chosen disturbance scenario. Thus, by considering higher disturbance values the advantage of the robust approach regarding stabilization would become even more notable. Naturally, this advantage is paid for by larger delay. However, in the robust version, the increase in delay amounts to the width of one time window w only. In our time window assignment approach, this is the smallest possible increase. The shape parameters of the robustified models exhibit near Poisson or even clustered characteristics $a < 1$ (mean stderr $< 5\%$) in contrast to baseline (Figure 6).

D. Discussion

The main goal of the present research was the inclusion of a-priori knowledge on disturbance statistics in the pre-tactical arrival sequence optimization through new stochastic and robust models and the validation of the increased scheduling stability. Practically relevant results were obtained by means of empirical arrival and departure delay statistics and a stochastic delay model for fitting the results of Monte Carlo (MC) computer experiments with 209 flights over 17 and 8 hours time span (low and high traffic scenarios S6.2, S7.2), and 200 repetitions each. The results are based on continuous time simulations (FCFS, TS8-2 optimizer) as baseline and on discrete optimization with 10 min TT-window using nominal (as additional baseline) and new stochastic and robust models. For each of the 200 repeated runs during an experiment random departure time delays $t^D(a_i)$ were drawn and added to the planned earliest and (max) latest times (ET(a_i), LTmax(a_i)).

Of course, FCFS without optimization provided the shortest runtime (< 2 s) but rather instable planning. However, more importantly the runtimes of the new MIP-models are also very low, namely 16 - 18 s / MC-run (note, that one MC-run contains around 150 simulation steps). These runtimes were significantly smaller than the continuous time TS8-2 baseline (> 200 s). Further, in the considered disturbance scenarios, the robust optimization approach needs almost no re-scheduling. Thus, it is by far the most stable approach, prior to the stochastic approach. This shows that it is indeed possible to stabilize pre-tactical planning by including knowledge about the uncertainties already in the modelling phase. The reasons are a better protection of the planning process against disturbance through reducing the effective ET $<$ LTmax range to ET $+\mu+\sigma <$ LTmax $+\mu-\sigma$ for each a_i of the sequence. The consequence of reduced overlap of disturbed effective arrival time intervals is obtained at the cost of the additional $+\sigma$ shift of each interval (LTmax$-$ET)$_i$ which in turn generates a corresponding delay-increase for MC089. In fact, the increase in delay amounts to the width of one time window only.

The difference between absolute delay values of baseline and MIP models is partly due to the fact that scheduling/optimization with TS8-2 and FCFS (continuous TT) was performed with regard to the TT$-$ET difference (due to requirement for monotonous objective function) whereas for the MIP-models the TT$-$ST difference and non-monotonous objective function was used. Consequently the delay levels of baseline exhibits a systematic deviation towards ET. On the other hand, also the large TT-window of the MIP-models generate a bias towards low delays due to the selection of the early edge of the windows as delay value for all a_i within w. Characteristically, robust and stochastic approaches seem to yield a Poisson-type (exponential, $a \leq 1$) delay PDF with standard deviation increased according to $cv = 1/\sqrt{a}$.

In order to evaluate the potential of the MIP-models in more detail the available parameters (k, μ, σ, TT-window w) and disturbance scenarios have to be modified in further computer experiments and different robustification variants can be tested. This also includes more advanced robustness concepts that reduce the potential conservatism. Such models are currently under development. As shown in [6], disturbances can have a significant influence on the nominal model in the sense that a considerable amount of reschedulings is necessary in order to make a solution feasible. In the scenarios considered here, the disturbances are less pronounced such that (in absolute values) the number of reschedulings is already quite small. For the future, it is thus interesting to validate our approaches also in scenarios with increased disturbances. It can be expected that the $rs\#$-reduction by using the robust optimization approach then will get pronounced even more. Further, for the present initial validation we used the same empirical average departure-delay PDF ($\langle\mu_i\rangle, \langle\sigma_i\rangle$) for all flights a_i. Within further validations the new models are expected to yield improved sequencing

TABLE III
COMPARISON OF NEW OPTIMIZATION MODELS WITH BASELINE

Model(MC#)	$\langle RT(\text{std})\rangle/s$	$\langle a\rangle$	$\langle\mu(\text{std})\rangle/\text{min}$	$\langle rs\#(\text{std})\rangle$
FCFS (57)	1.1 (0.4)	2.5	5.3 (5.0)	12.7 (5.6)
TS8-2 (56)	203 (63)	3.5	3.5 (6.0)	8.1 (3.2)
Nominal (87)	15.8 (1.5)	3.2	1.7 (3.9)	0.45 (0.27)
Stochastic (88)	16.7 (1.1)	0.9	3.0 (5.1)	0.38 (0.23)
Robust (89)	18.2 (1.2)	0.3	12.2 (0.6)	0.22 (0.26)

results through individualized disturbance and ET- and LT-shift values (μ_i, σ_i) derived from the single-flight PDF's.

V. CONCLUSION

In our mathematical model for pre-tactical planning, several aircraft can be assigned to the same time window which reduces the complexity of the problem. Details are described in a previous publication ([6]). We enriched this model by protection against uncertainties using techniques from robust and stochastic optimization.

Initial validation of the new models was performed by means of Monte Carlo (MC) computer experiments. For deriving a departure delay model to generate realistic disturbances for the MC simulations we performed a statistical analysis of real-world data from a large German airport. Furthermore, we described the simulation environment for these experiments in order to validate the different optimization approaches. The data analysis together with the baseline simulations indicate the two-parametric Γ-PDF to be a reasonable approach for deriving stochastic performance metrics. The scheduling performance of the new MIP-models with stochastic and robust protection against disturbance were quantified with regard to runtime, re-scheduling stability and arrival delay statistics (shape and mean value). Compared with baseline scheduling they exhibit the predicted significantly reduced runtime and re-scheduling, to be paid for by an increase of delays. However, this delay is at most the width of about one time window. Furthermore they exhibit more exponential than skewed-Gaussian like distributions. The stochastic approach optimizes the expected scenario and, therefore, is more likely to remain feasible than the nominal approach. It however is less likely to be feasible than the robust approach. Using the robust approach, we definitely know that a solution will be feasible for all scenarios within the pre-determined uncertainty set. Thus, it is the approach with the highest possible stability. The initial and preliminary validation results need confirmation by additional computer experiments, which are ongoing. We will also include and validate more advanced robustness concepts with reduced conservatism that are already in development. However, the encouraging results already show that we succeeded in computing stable plans with a high probability to remain feasible despite changes in the input data.

ACKNOWLEDGMENT

This work is co-financed by EUROCONTROL acting on behalf of the SESAR Joint Undertaking (the SJU) and the EUROPEAN UNION as part of Work Package E in the SESAR Programme. Opinions expressed in this work reflect the authors' views only and EUROCONTROL and/or the SJU shall not be considered liable for them or for any use that may be made of the information contained herein.

We are grateful to A. Martin for many stimulating discussions about the optimization aspects of the topic addressed here. Further, we are indebted to Olga Gluchshenko who prepared the initial dataset for the MC-simulations. Many thanks are due to S. Loth for paving the way to the airport data sources and to M. Helms for initial preparation of the large amount of empirical data from that source. We would like to thank H. Helmke for valuable advice throughout this work concerning the optimization aspects and to J. Rataj for continuous support.

REFERENCES

[1] Abdel-Aty, M., Lee, C., Bai, Y., Li, X., Michalak, M. (2007). *Detecting periodic patterns of arrival delay*. Journal of Air Transport Management, vol. 13, S. 355-361.

[2] Ben-Tal, A., El Ghaoui, L., Nemirovski, A. (2009). *Robust Optimization*. Princeton University Press, Princeton.

[3] Bertsimas, D., Sim, M. (2003). *Robust Discrete Optimization and Network Flows*. Mathematical Programming, 98:49-71.

[4] Caccavale, M. V., Iovanella, A., Lancia, C., Lulli, G., & Scoppola, B. (2014). *A model of inbound air traffic: The application to Heathrow airport*. Journal of Air Transportation Management, vol. 34. S. 116-122.

[5] Dodson, C., Scharcanski, J. (2003). *Information Geometric Similarity Measurement for Near-Random Stochastic Processes*. IEEE Trans. Syst., Man, Cybernetics - A: Systems and Humans, vol.33, no.4, S. 435-440.

[6] Fürstenau, N., Heidt, A., Helmke, H., Kapolke, M., and Liers, F. (2014). *Pre-Tactical Time Window Assignment: Runway Utilization and the Impact of Uncertainties*. Schaefer, Dirk (Editor) Proceedings of the SESAR Innovation Days (2014) EUROCONTROL. ISBN 978-2-87497-077-1.

[7] Guadagni, G., Ndreca, S., & Scoppola, B. (2011). *Queuing systems with pre-schduled random arrivals*. Math. Methods of Operations Research, vol. 73, S. 1-18.

[8] Helmke, H. (2011). *Scheduling algorithms for ATM applications - tools and toys*. Proc. 30th Digital Avionics Systems Conf. Seattle.

[9] ICAO Document 4444 (2007). *Procedures for air trafic management*. 15th edition, pp. 69-126.

[10] Kall, P., Mayer, J. (2013). *Stochastic Linear Programming: Models, Theory and Application*. Springer, Wiesbaden.

[11] Performance Review Commission. (2013). *Performance Review Report PRR2012*. Brussels: Eurocontrol.

[12] Tu, Y., Ball, M., Jank, W. (2008). *Estimating Flight departure delay distributions - a statistical approach with long-term trend and short-term pattern*. J. of the American Statistical Association, vol. 103, S. 112-125.

[13] Wong, J.-T., Tsai, S.-C. (2012). *A survival model for flight delay propagation*. Journal of Air Transport Management, vol. 23, S. 5-11.

[14] Wu, C.-L. (2010). *Airline operations and delay management*. Farnham / GB: Ashgate.