Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)

Analysis on the Impact of Pop-Up Flight Occurrence when Extending the Arrival Management Horizon

A. Vanwelsenaere, J. Ellerbroek, and J. M. Hoekstra
Control and Simulation, Faculty of Aerospace Engineering
Delft University of Technology (TU Delft)
Delft, The Netherlands

E. Westerveld
Air Traffic Control The Netherlands (LVNL)
Schiphol, The Netherlands

Abstract

The occurrence of pop-up flights negatively affects the (extended) arrival manager. This issue is known already for a long time by operational experts, but the extent thereof has now been assessed during experiments. An arrival manager research model was developed and integrated in BlueSky, an open-source air traffic management simulator. Fast-time simulations showed that extended arrival management is significantly negatively affected by pop-up flights, in terms of flight crew and air traffic control task load, sequence stability and delay (cost). Simulations also indicated that this impact could be mitigated by pre-planning pop-up flights prior to departure, using their take-off time estimates. This will, however, only be beneficial when these estimates are sufficiently accurate. With currently achievable accuracies, and using currently available systems, it is better to discard these estimates in the context of extended arrival management.

I. INTRODUCTION

In Europe, air traffic is concentrated on a relatively limited number of major hub airports. Flights to these airports need to absorb delays prior to landing, as these airports experience short-term capacity-demand imbalances. Inbound traffic is guided from upper airspace en-route sectors towards the destination airport. Delays can be absorbed through speed reductions, vectoring, or by placing aircraft in holding stacks. [1].

Europe's busiest airports have implemented arrival management (AMAN) systems to mitigate short-term capacity-demand imbalances. Because of the decentralized nature of European development of AMAN systems, as well as the airspace design, there is a large variation in the actual working principles and usage between systems at different airports [2], [3]. What these airports have in common, however, is the desire to have aircraft absorb more of their delays en-route, as this increases operational efficiency. [4]

The involved air navigation service providers are therefore currently examining a horizon extension of their AMAN systems, referred to as the extended arrival management (E-AMAN) concept. With this increased horizon, delays can be absorbed more efficiently upstream. Two issues, however, arise with the introduction of an extended AMAN horizon: inaccuracies related to trajectory prediction (TP), and the occurrence of pop-up flights [5]–[7]. While over the last decades, various studies have been published on analysis and improvement of the TP process [8]–[12], only little research has been done on the occurrence of pop-up traffic.

Pop-up flights are flights that depart within the horizon of the AMAN system, implying that these flights still need to join the arriving traffic stream when the sequence has already been established. Often the schedule needs to be revised, which could seriously disrupt the arrival management process. Even though pop-up flights - also referred to as in-horizon departures - pose one of the most significant operational and technical difficulties [3], little research has been performed on the occurrence and effect thereof. Related studies, however, indicate that inaccuracies related to the Estimated Time of Arrival (ETA) are substantially larger when the aircraft has not departed yet (which is the case with in-horizon departures); ETAs for airborne flights, even at large distances from the airport, are significantly more accurate when compared to those of pop-up flights. [13], [14]. When compared to TP errors, the negative impact of pop-up flights on the arrival management process is therefore considered substantially larger. Due to the horizon extension of E-AMAN, the number of in-horizon departures increases even more. Consequently their negative impact might grow as well.

The work presented in this paper will examine the effects of pop-up flights on E-AMAN, and will evaluate mitigating measures. Two experiments were set up. In the first experiment, the actual effect of pop-up occurrence on E-AMAN was assessed. It was analysed to which extent pop-up occurrence would hinder the implementation of E-AMAN, if no action were taken when extending the AMAN horizon. Preliminary results indicated that the negative impact of pop-up flights is large, and therefore measures need to be taken. Therefore a second experiment was set up, in which pop-up flights were pre-planned based on their estimated take-off time. Results of both experiments were analysed, after which conclusions and recommendations were formulated. The remainder of this paper is structured as follows: Section II describes the AMAN research model that was used during the fast-time simulations. Section III outlines the first experiment, in which the actual impact of pop-up flights on E-AMAN has been analysed. The second experiment, to assess the effect of pre-planning pop-up flights before departure, is described in section IV. The paper ends with a discussion and conclusion on the project outcomes.

II. AMAN RESEARCH MODEL

For the experiments, an AMAN research model was developed and integrated in BlueSky, an open-source ATM simulator being developed at Delft University of Technology [15]. This study is focused on the AMAN system of Amsterdam Schiphol Airport (ICAO: EHAM). The working principles of the AMAN research model are therefore based on those of the Advanced Schiphol Arrival Planner (ASAP); a new AMAN that is currently under deployment. Certain advanced features were omitted or simplified in order to reduce the model's development time.

The basic working principle of ASAP is as follows: once radar data is available, the trajectory predictor (TP) periodically derives the Estimated Time of Arrival for that aircraft. Based on the ETAs of all flights, the scheduler sets up a schedule and assigns a Scheduled Time of Arrival (STA) to all applicable aircraft. Delays are supposed to be absorbed prior to the Terminal Maneuvering Area (TMA) entry, hence STAs are translated to Expected Approach Times (EATs) at the Initial Approach Fix (IAF). In practice, this means that en-route controllers (Area Control) should hand over aircraft to approach controllers at the IAF around this expected time.

The research model periodically derives the ETAs for all aircraft within the AMAN eligibility horizon (EH), i.e., the horizon from which radar data is available. TP inaccuracies have been reduced pro-actively to a minimum, since these errors might otherwise confound the effects of pop-up occurrence.

The ASAP scheduler gathers the ETAs for all applicable flights, sets up a planning, and assigns STAs. For aircraft outside the AMAN Freeze Horizon (FH), which is typically substantially smaller than the EH, the schedule and corresponding STAs are updated and revised continuously using a First Come First Served (FCFS) algorithm. Once flights enter the FH, their STA is in principle fixed, unless a pop-up flight departs and triggers a schedule revision. Pop-up flights are those aircraft that depart within the FH, and possibly impose STA revisions to (multiple) airborne flights. In the framework of AMAN and E-AMAN, the FH is set to 120 nm and 200 nm respectively in the model [7], [16]. In the scheduler, pop-up flights are only considered once airborne, as is currently the case for most European AMANs.

In ASAP, Area Control is responsible for delivering aircraft at the IAF around their EAT (margin of ±30 seconds). Once aircraft enter the active advisory horizon (AAH), which is typically slightly smaller than the FH, Area Control can provide commands in order to match the desired arrival time. To assist controllers in generating the necessary advisories that deliver the aircraft within the tight time constraints, the Speed and Route Advisor (SARA) tool has been integrated in ASAP [2]. A simplified SARA module was developed for the research model, which automatically generates flight-plan revisions to deliver aircraft around the EAT at the IAF. Speed advisories can be provided by means of reducing the aircraft's speed up to 10% over the remaining trajectory. Route advisories are provided when the delays are too large to be absorbed with speed reductions only. This route advisory generation process has been simplified by means of concentrating the advisories on the last leg prior to the TMA entry. Depending on the magnitude of the delays to be absorbed, aircraft are vectored or placed in so-called holding stacks.

III. EXPERIMENT I: THE EFFECT OF POP-UP FLIGHTS

An experiment was performed to assess how pop-up flights impact the (extended) arrival management process, by observing several ratios of pop-up flights in the samples. In addition, the implications of a horizon extension from AMAN to E-AMAN are assessed.

A. Experiment Set-Up

1) Apparatus and Model

Simulations have been performed using BlueSky [15], with the AMAN research model presented in Section II, to assess the effect of pop-up flights. Simulations were carried out in fast-time, where controllers were assumed to always follow the SARA advisories presented in Section II. Conflict detection and resolution functions were disabled, as they could interfere with the arrival management process. In reality, ATC uses AMAN systems as a complementary tool, and performs the tactical functions simultaneously. All simulations were performed for Amsterdam Schiphol Airport, see Fig. 1. Runways 18C and 27 were used for arrivals, each with a landing interval of 100 seconds. Based on the flight plan's IAF, runways were allocated during the scenario generation process. Here, the SUGOL IAF fed runway 18C, ARTIP fed runway 27, and RIVER fed both randomly.

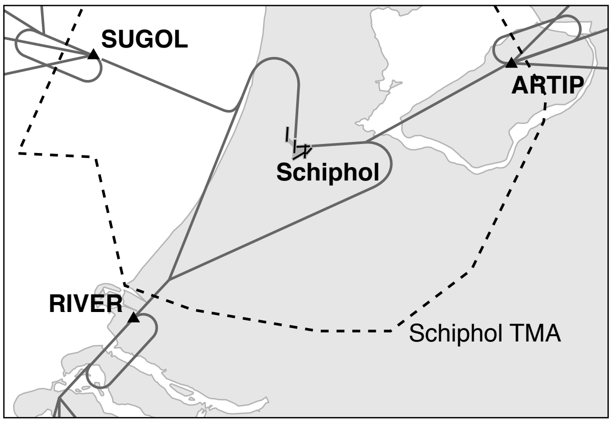

Fig. 1. Approach routes in simulation.

2) Independent Variables

This experiment investigates the effect of pop-up flights on arrival management for both the existing AMAN, and the envisioned E-AMAN systems. The AMAN *Freeze Horizon* (FH) is therefore an independent variable with two levels: FH could be either 120 nm (AMAN), or 200 nm (E-AMAN). This implies that in the E-AMAN situation, 11.5% of all arriving flights concerns pop-up traffic. In the AMAN situation, this is evidently lower (1.9%). During the fast-time simulations, the actual occurrence of pop-up flights was tuned to assess the effect thereof. *Pop-up Scaling* (PS) is defined as the scaling of the relative occurrence of pop-up flights in percent. If PS is 100%, pop-up flights occur as in the original sample (for the applicable horizon). This implies that approximately 1.9% and 11.5% of all arriving flights, in the AMAN and E-AMAN situation respectively, concerns pop-up flights if $PS = 100\%$. If PS is smaller or larger, pop-up flights have been replaced by longer-haul flights (or vice versa). By doing so, the actual traffic demand remains similar and the outcomes of the experiment conditions could be compared. In the experiment, pop-up scaling is an independent variable with three levels; respectively $PS = 0\%$, $PS = 100\%$, and $PS = 200\%$. The experimental conditions are summarized in Table I.

3) Experiment Design and Traffic Samples

The experiment was designed as a within-subjects, repeated-measures, where experimental conditions were compared using twelve different traffic samples. These samples consist of ETFMS flight plan data[1] for a given time window, complemented with ASAP's TMA routes (section II). If the occur-

---

[1] Called M2 trajectories in the Eurocontrol Demand Data Repository.

TABLE I
CONDITIONS[†] FOR EXPERIMENT I

Exp. Condition	AMAN FH	Pop-Up Scaling	Pop-Up Occurrence
A/0	120 nm	0%	0.0%
A/100		100%	1.9%
A/200		200%	3.8%
E/0	200 nm	0%	0.0%
E/100		100%	11.5%
E/200		200%	23.0%

[†] The first and last three conditions correspond to the AMAN and E-AMAN context, respectively.

rence of pop-up flights was altered by replacing pop-up flights with longer-haul flights (or vice versa), items in the original sample were modified. Samples were drawn from peak periods on six weekdays during the summer of 2015; each sample concerns inbound traffic for a three-hour period. The actual traffic in each sample varies, ranging from 128 up to 145 flights.

4) Dependent Measures

The overall performance of the (extended) arrival management process was evaluated in terms of delay (cost), sequence stability, and task load of flight crew and air traffic control. *Delay (cost)* was measured in two ways: both in terms of the average required delay absorption due to low-level route advisories (LLDA: low-level delay absorption), and in terms of the average energy cost caused by absorbing airborne delays, normalized by flight-plan distance. Here, energy was derived using a simple energy cost model based on equations of motion, and assuming steady and quasi-linear flight.

The number of Scheduled Time of Arrival (STA) revisions, and the number of disturbed descents were used as measures of *task load*. Here, a descent was considered disturbed when the STA of an aircraft is revised after Top of Descent (ToD). The ToD is typically located approximately 100-120 nm away from the airport. Finally, the number of arrival sequence position changes was used to assess *sequence stability*; this measure is referred to as position changes in the remainder of the paper.

B. Hypotheses

Pop-up flights are expected to disturb the extended arrival management (E-AMAN) process. It is therefore hypothesized that a higher occurrence of pop-up flights negatively affects the following dependent measures: delay (cost) (H1-1), sequence stability (H1-2), flight crew and ATC task load (H1-3). As the pop-up flights occur significantly less frequent within the AMAN horizon, it is hypothesized that the negative effect of pop-up scaling is smaller for AMAN compared to E-AMAN (H1-4). Moving from AMAN to E-AMAN for actual pop-up occurrence levels is hypothesized to have a positive impact on delay (cost) (H1-5). However it is hypothesized to have a negative effect on sequence stability (H1-6) and on task loads (H1-7).

C. Results

Even though all samples are relatively similar in terms of overall traffic demand, the simulation outcomes deviate substantially from sample to sample within a given experiment condition. This can be attributed to the fact that the actual demand evolution is different in every sample. In addition, the effect of pop-up flights might be larger or smaller in a particular sample, depending on when and where the pop-up flights departed. For these reasons, statistical analyses were carried out using normalized and standardized Z-scores. The magnitude of each effect was analysed using the average of the results. Table II shows these averages for all E-AMAN conditions (i.e., for the three levels of pop-up scaling), as well as for condition A/100 (AMAN situation with PS=100%).

Shapiro-Wilk tests indicated that the majority of the Z-score distributions are not normally distributed. Therefore, only non-parametric tests were used for statistical analysis. Pop-up scaling was considered as a main effect for the AMAN and E-AMAN case separately, using Friedman's ANOVA test. Effects were considered significant for $p \leq 0.05$. Post-hoc tests were performed using a Wilcoxon's Signed Rank test. Both for the AMAN and the E-AMAN case, two post-hoc tests were performed, which compared the nominal condition ($PS = 100\%$) to $PS = 0\%$ and $PS = 200\%$, respectively. In addition, the effect of a horizon extension was considered at normal pop-up scaling (A/100 and E/100). With five pairs in total, a Bonferroni correction of 5 is used. Hence, post-hoc tests were considered significant when $p \leq 0.01$.

TABLE II
EXPERIMENT I: SAMPLE AVERAGES.

Exp. Condition	A/100	E/0	E/100	E/200
Av. LLDA per ac [s]	140.95	119.12	116.10	117.71
Delay e. c. [MJ/nm]	27.53	20.74	21.52	23.68
Position changes	6.33	26.17	40.83	70.83
STA revisions	0.92	8.92	24.83	43.50
Disturbed descents	12.08	0.17	0.75	1.58

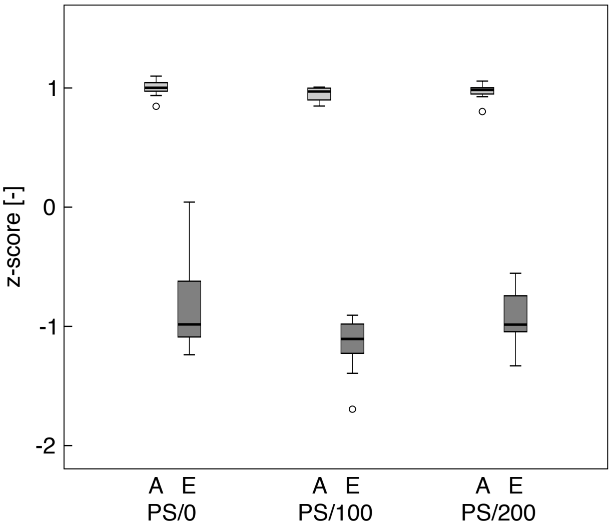

Fig. 2. Average low-level delay absorption (A=AMAN, E=E-AMAN, PS=Pop-up Scaling).

1) Delay (Cost)

Fig. 2 shows the results for the average low-level delay absorption. For both AMAN (A) and E-AMAN (E), the results are shown for the three pop-up scaling options: PS/0 where pop-up flights are removed in the original sample and replaced by longer-haul flights, PS/100 with pop-up occurrence as in the original sample, and PS/200 where pop-up flight occurrence was doubled by replacing longer-haul flights in the original sample by pop-up traffic. For both AMAN and E-AMAN, no clear effect of pop-up scaling can be seen. A main effects test also showed no significant effects of pop-up scaling ($\chi^2(2) = 2.00, p = 0.37$). Runways have a given capacity, based on the inter-arrival time. When the demand nears or exceeds capacity, aircraft need to absorb the necessary delays. The occurrence of pop-up flights does not alter the ratio between demand and capacity. It therefore makes sense that the required degree of low-level delay absorption is not affected by pop-up scaling.

A post-hoc comparison of condition E/0 and E/100 revealed that significantly fewer aircraft required low-level delay absorption in the E/100 case ($z = 2.98, p < 0.01$). This seems counter-intuitive, however, it can be argued that schedule revisions are beneficial in this perspective. When re-scheduling aircraft as they are closer to the runway, trajectory prediction errors are reduced, and the (new) schedule is set up using fewer uncertainties. In reality, TP errors significantly affect AMAN efficiency; these TP errors have been reduced substantially in the research model, however they could not be eliminated. When doubling the pop-up occurrence (E/200) with respect to E/100, no significant effect of pop-up scaling was identified ($z = 1.58, p = 0.11$).

The occurrence of pop-up flights is therefore, to some extent, beneficial in terms of reducing the number of aircraft that require delay absorption. On average, 7% less flights require delay absorption when comparing E/100 with E/0. However, when pop-up scaling is increased (E/200), there is no additional benefit when compared to E/100. For AMAN, none of the post-hoc tests revealed a significant difference. This is attributed to the fact that the occurrence of pop-up flights (on average 1.9% in A/100) is too small to observe certain effects. In addition, due to this low occurrence in AMAN, the actual magnitude of the effects is small. Because of this, the remainder of the analysis in this paper will focus primarily on the context of E-AMAN.

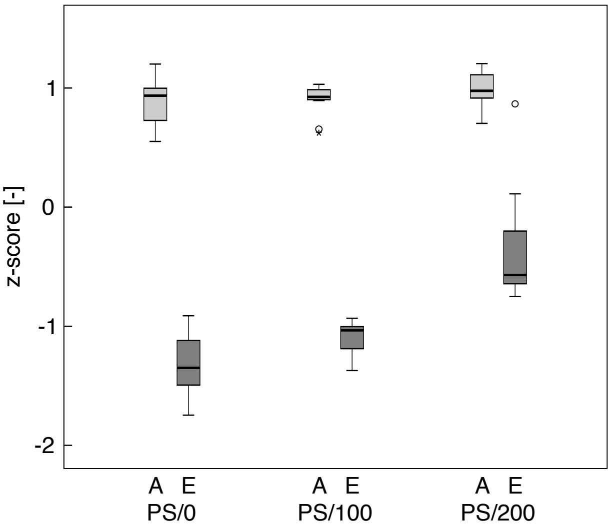

Fig. 3. Delay energy cost (A=AMAN, E=E-AMAN, PS=Pop-up Scaling).

The delay energy cost is shown in Fig. 3. For both AMAN (A) and E-AMAN (E), the results are shown for the three pop-up scaling options: PS/0 where pop-up flights are removed in the original sample and replaced by longer-haul flights, PS/100 with pop-up occurrence as in the original sample, and PS/200 where pop-up flight occurrence was doubled by replacing longer-haul flights in the original sample by pop-up traffic. It can be seen that the delay energy cost increases with increasing pop-up scaling. This effect was significant for E-AMAN ($\chi^2(2) = 19.50, p < 0.01$). Post-hoc tests show that E/200 differs significantly from E/100 ($z = 3.06, p < 0.01$), where cost is 10% larger on average in E/200, see also Table II. While post-hoc tests did not indicate a significant difference between E/100 and E/0 ($z = 1.80, p = 0.07$), there is a tendency of increased cost for larger pop-up occurrences, as shown in Fig. 3. The overall trend that can be observed from the results, however, is that the larger the uncertainties are, as induced by the pop-up flights, the higher their negative effect on energy cost. Similar to the required delay absorption results, no significant effects were found for pop-up scaling in the AMAN case.

As the horizon is extended, more delays can be absorbed by en-route speed reduction. The disturbances, induced by the increased number of pop-up flights in E-AMAN, are outweighed by the benefits of increased delay absorption. By comparing E/100 with A/100, the average low-level delay absorption can be reduced by 17%; post-hoc tests indicate that this result is significant ($z = 3.06, p < 0.01$). As more delays can be absorbed en-route, 26% fewer aircraft require low-level delay absorption ($z = 3.06, p < 0.01$, comparing E/100 to A/100). In this comparison, delay energy cost is reduced by 22% due to the extended horizon. This difference was also significant ($z = 3.06, p < 0.01$). These are averaged results, implying that the actual effect magnitude varied depending on the sample. Nevertheless, comparable trends were observed for all traffic samples.

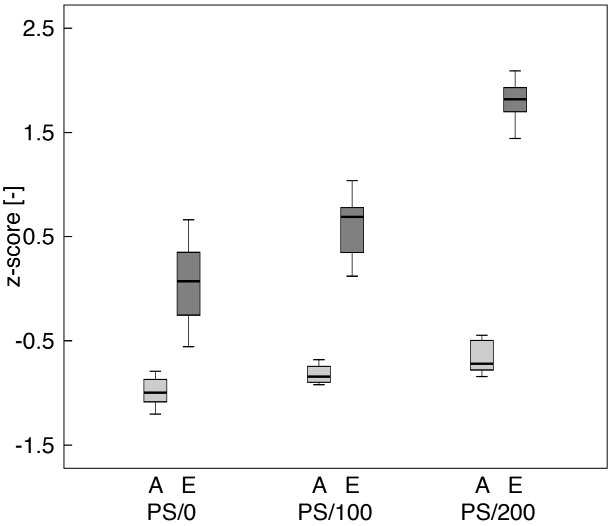

Fig. 4. Arrival sequence position changes (A=AMAN, E=E-AMAN, PS=Pop-up Scaling).

2) Sequence Stability

The results in terms of arrival sequence position changes, an indicator for sequence stability, are shown in Fig. 4. For both AMAN (A) and E-AMAN (E), the results are shown for the three pop-up scaling options: PS/0 where pop-up flights are removed in the original sample and replaced by longer-haul flights, PS/100 with pop-up occurrence as in the original sample, and PS/200 where pop-up flight occurrence was doubled by replacing longer-haul flights in the original sample by pop-up traffic. Here it can be seen that both the increased pop-up scaling, as well as the increased planning horizon, result in an increased number of position changes. A main effects test revealed that the the influence of pop-up scaling was significant in the E-AMAN conditions ($\chi^2(2) = 22.17, p < 0.01$). Post-hoc tests showed that E/100 differs significantly from E/0 ($z = 2.82, p < 0.01$), and that E/200 differs significantly from E/100 ($z = 3.06, p < 0.01$). The sample averages of the three E-AMAN conditions (Table II) show that this effect is large: with respect to E/0, the required number of position changes in E/100 increased by 56% on average. When comparing E/200 with E/100, this is increased by an additional 75%. The absolute values show the large negative impact of pop-up occurrence on sequence stability experienced in the context of E-AMAN. Similar statistical results were found in the framework of AMAN, although the actual negative effect is negligible due to the low pop-up occurrence in all AMAN conditions.

Post-hoc analysis also shows that the effect of the increased planning horizon was significant, when comparing conditions A/100 and E/100 ($z = 3.06, p < 0.01$). On average, the required number of position changes is 6 times larger in E/100 when compared to A/100, see Table II.

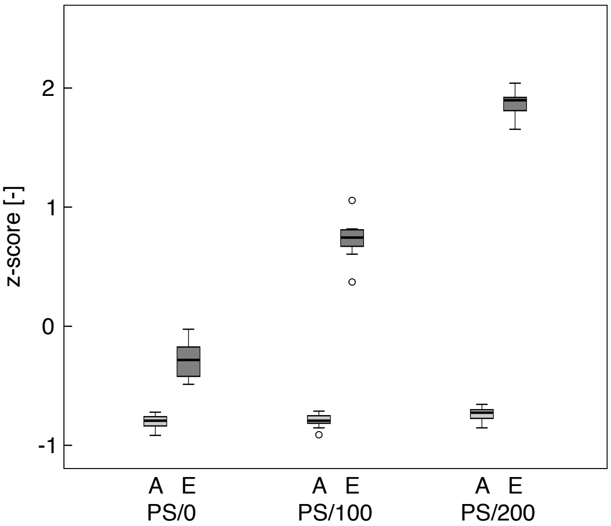

Fig. 5. STA revisions (A=AMAN, E=E-AMAN, PS=Pop-up Scaling).

3) *Flight Crew and ATC Task Load*

Fig. 5 shows the results in terms of the number of STA revisions. For both AMAN (A) and E-AMAN (E), the results are shown for the three pop-up scaling options: PS/0 where pop-up flights are removed in the original sample and replaced by longer-haul flights, PS/100 with pop-up occurrence as in the original sample, and PS/200 where pop-up flight occurrence was doubled by replacing longer-haul flights in the original sample by pop-up traffic. Similar to the previous metric, it can be seen that increased pop-up scaling has a negative effect on the required number of STA revisions, which was significant for both AMAN ($\chi^2(2) = 13.00, p < 0.01$), and E-AMAN ($\chi^2(2) = 24.00, p < 0.01$). Post-hoc tests showed a significant difference between E/0 and E/100 ($z = 3.06, p < 0.01$), and between E/100 and E/200 ($z = 3.06, p < 0.01$). The disturbances grow as the pop-up occurrence increases. By comparing E/0 and E/100, both variables increased by nearly factor 3; in E/200, on average 75% more STA revisions occur when compared to E/100. Both in relative and absolute terms, these effects are large. It should be realized that the occurrence of pop-up flights increases from 11.5% (E/100) to 23.0% (E/200). For AMAN, post-hoc tests did not reveal significant differences between pairs ($p > 0.01$), which is attributed to the low occurrence of pop-up flights with AMAN.

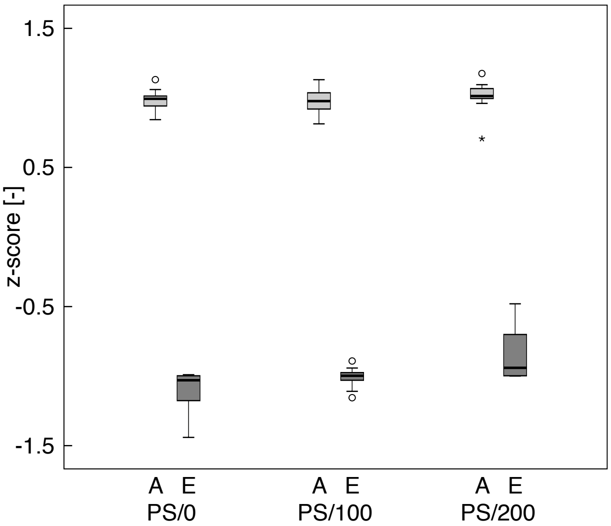

Fig. 6. Disturbed descents (A=AMAN, E=E-AMAN, PS=Pop-up Scaling).

The results in terms of the number of disturbed descents are shown in Fig. 6. For both AMAN (A) and E-AMAN (E), the results are shown for the three pop-up scaling options: PS/0 where pop-up flights are removed in the original sample and replaced by longer-haul flights, PS/100 with pop-up occurrence as in the original sample, and PS/200 where pop-up flight occurrence was doubled by replacing longer-haul flights in the original sample by pop-up traffic. A main effects test revealed that the effect of pop-up scaling is significant for E-AMAN ($\chi^2(2) = 10.30$, $p < 0.01$), but not for AMAN ($\chi^2(2) = 2.85, p = 0.24$). Post-hoc tests of the E-AMAN results, however, did not reveal significance between pairs ($p > 0.01$). Nevertheless, Fig. 6 does show a tendency of more disturbed descents for increased pop-up occurrence. The sample averages are close to zero for all conditions (Table II), implying that the effect, even if it would be significant, is very small. While this may seem counter-intuitive, it can be explained by the fact that most pop-up aircraft depart prior to the Top of Descent (ToD) of airborne aircraft. Airborne aircraft therefore rarely experience disturbed descents due to the occurrence of pop-up flights.

Post-hoc comparisons of A/100 and E/100 show that the effects of the horizon extension were significant both for the number of STA revisions ($z = 3.06, p < 0.01$), and for the number of disturbed descents ($z = 3.06, p < 0.01$). In Table II it can be seen that the number of STA revisions is, on average, 27 times larger with E-AMAN. This is caused by the higher pop-up occurrence, which is, on average, 6 times higher in E/100. The number of disturbed descents is reduced to nearly zero in E/100, whereas its occurrence (9% on average) in AMAN is not considered problematic either.

IV. EXPERIMENT II: ALTERNATIVE SCHEDULER

The results of the first experiment show that pop-up flights negatively affect the (extended) arrival management process. This effect might be mitigated by taking pop-up flights into account, prior to their departure. A second experiment was therefore performed to assess the benefit of pre-planning pop-up flights prior to their departure. This experiment evaluates an alternative scheduler that takes this into account, for various levels of accuracy of the departure time estimate.

A. Experiment Set-Up

1) Apparatus and Model

Similar to the first experiment, fast-time simulations were performed with the AMAN research model. All runs simulated arrivals to EHAM, using the landing interval and runway allocation procedure as applied previously. Compared to the previous experiment, this experiment considers only the extended AMAN horizon, with pop-up occurrence as in current traffic ($PS = 100\%$). To ensure that conditions are comparable, departure information accuracy of all pop-up flights was constant within each experiment run.

2) Independent Variables

To assess the effect of pre-planning pop-up flights prior to departure, an alternative scheduler was used that takes this into account. It explicitly uses the pre-departure take-off time estimates of pop-up flights to plan them along with the airborne aircraft. When the pop-up flight departs at its estimated time, no substantial schedule revisions are required. However, if the pop-up flight departs earlier or later, its reserved place in the sequence needs to be revised once the aircraft gets airborne, possibly impacting airborne aircraft. *Pre-planning* was therefore an independent variable, with five levels: pre-planning could be either absent (i.e., the original scheduler is used), or pre-planning was applied with departure estimate uncertainties U/0 (pop-up aircraft departs exactly at its estimated time), U/120 (departs 2 minutes - or 120 seconds - later), U/180 (departs 3 minutes - or 180 seconds - later) and U/300 (departs 5 minutes - or 300 seconds - later). The experimental conditions are summarized in Table III.

TABLE III
CONDITIONS FOR EXPERIMENT II

Exp. Condition	**Scheduler**	**Dep. Uncertainty**
Baseline	no pre-planning	NA†
U/0	pre-planning	0 s
U/120		120 s
U/180		180 s
U/300		300 s

† Not applicable, as the baseline scheduler does not pre-plan pop-up flights prior to departure.

3) Experiment Design and Dependent Measures

Similar to the first experiment, the second experiment was designed as a within-subjects, repeated measures. The same traffic samples were used to compare the experimental conditions. Also the same dependent measures were used to assess the effect of pre-planning.

B. Hypotheses

It is hypothesized that the pre-planning scheduler outperforms the baseline scheduler when the take-off time estimates are perfect (H2-1). In a previous study, Barnier and Allignol found that for aircraft deconfliction, incorporating flights prior to departure was not effective with departure time uncertainties of three minutes [17]. It was therefore hypothesized that for pre-planning of pop-up flights to be effective, the departure time uncertainty needs to be smaller than three minutes (H2-2).

C. Results

The statistical analysis process is similar to Experiment I. Z-scores were used to assess the results. Shapiro-Wilk tests on the data revealed that for the majority of the data, normality could not be assumed. Friedman's ANOVA was therefore used to evaluate the main effect, where effects were considered significant for $p \leq 0.05$. The Wilcoxon's Signed Rank test was used as a post-hoc test for five pairs: the baseline condition compared to the other four conditions, and U/0 compared to U/120. Using a Bonferroni correction of 5, post-hoc tests were considered significant when $p \leq 0.01$.

TABLE IV
EXPERIMENT II: SAMPLE AVERAGES.

Exp. Condition	Basic	U/0	U/120	U/180	U/300
Av. LLDA per ac [s]	116.1	116.0	121.5	128.1	128.9
Delay e. c. [MJ/nm]	21.5	21.4	22.0	22.3	22.4
Position changes	40.8	48.1	61.00	61.5	73.1
STA revisions	24.8	15.3	27.8	31.1	36.2
Disturbed descents	0.8	0.8	1.7	2.0	3.1

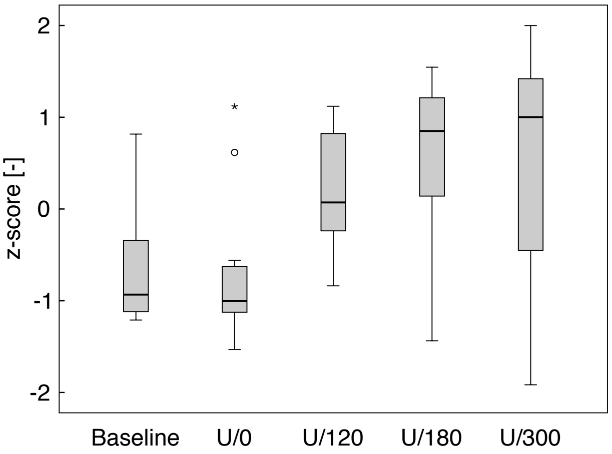

Fig. 7. Average low-level delay absorption (U=Uncertainty Departure Time Estimate).

1) Delay (Cost)

Fig. 7 shows the results in terms of low-level delay absorption. In this figure, the result of the baseline situation (no pre-planning of pop-up flights) is shown along with the results of the four conditions in which pre-planning of pop-up traffic takes place: U/0 (pop-up aircraft departs exactly at its estimated time), U/120 (departs 2 minutes later), U/180 (departs 3 minutes later) and U/300 (departs 5 minutes later). A main effect test revealed a significant effect of pre-planning ($\chi^2(4) = 13.93, p < 0.01$). Post-hoc tests revealed that scheduling pop-up flights with perfect accuracy does not affect the average low-level delay absorption, as condition U/0 did not differ significantly from the baseline ($z = 0.00, p = 1.00$). As long as the demand-capacity ratio of the runway is not substantially altered, the required delay absorption remains similar. When increasing the take-off time estimate error, however, delay absorption increases, resulting in a significant difference between U/120 and U/0 ($z = 2.67, p < 0.01$). This can be attributed to the fact that with pre-planning errors, more aircraft need to absorb larger delays, are informed about this at a late stage and therefore require inefficient delay absorption at low altitude. Compared to the baseline, however, none of the degraded estimate conditions showed a significant difference ($p > 0.01$).

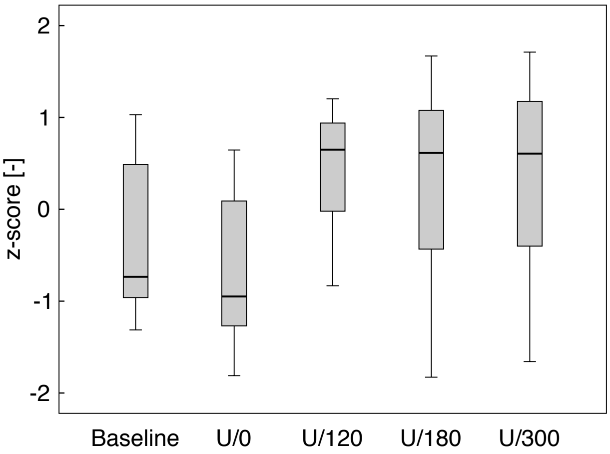

Fig. 8. Delay energy cost (U=Uncertainty Departure Time Estimate).

Delay energy cost is shown in Fig. 8. In this figure, the result of the baseline situation (no pre-planning of pop-up flights) is shown along with the results of the four conditions in which pre-planning of pop-up traffic takes place: U/0 (pop-up aircraft departs exactly at its estimated time), U/120 (departs 2 minutes later), U/180 (departs 3 minutes later) and U/300 (departs 5 minutes later). A main effect was observed ($\chi^2(4) = 9.40, p < 0.05$), but post-hoc tests only revealed a significant difference between conditions U/0 and U/120 ($z = 2.98, p < 0.01$). Once take-off time estimate accuracies deteriorate, cost increases. For the conditions in which estimate errors were included in the pre-planning scheduler, no statistically significant differences were identified when compared to the baseline condition. Fig. 8, however, does illustrate a tendency of growing energy cost when inaccuracies are included in the pre-planning scheduler.

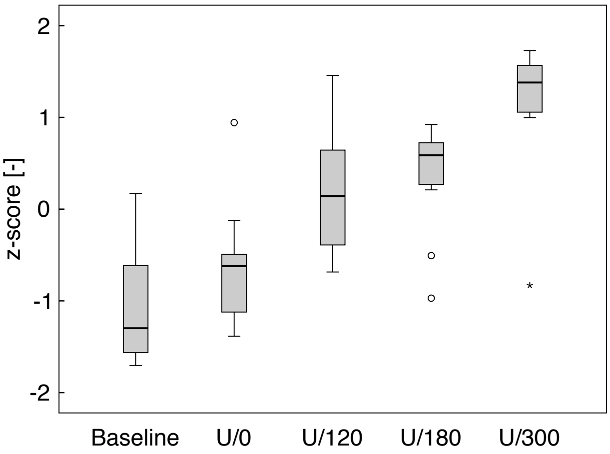

Fig. 9. Arrival sequence position changes (U=Uncertainty Departure Time Estimate).

2) *Sequence Stability*

Fig. 9 shows the sequence stability results in terms of the number of arrival sequence position changes. In this figure, the result of the baseline situation (no pre-planning of pop-up flights) is shown along with the results of the four conditions in which pre-planning of pop-up traffic takes place: U/0 (pop-up aircraft departs exactly at its estimated time), U/120 (departs 2 minutes later), U/180 (departs 3 minutes later) and U/300 (departs 5 minutes later). A main effects test revealed a significant influence of pre-planning on the number of position changes ($\chi^2(4) = 32.82, p < 0.01$). Post-hoc tests revealed no significant effect of pre-planning with accurate estimates (U/0), compared to the baseline ($z = 1.57, p = 0.12$). However, compared to the baseline the number of position changes increases significantly for all of the deteriorated estimate conditions ($p < 0.01$). Average statistics of deteriorated condition U/300, for instance, show an increase of 79% in the number of position changes, when compared to the baseline (see Table IV). Once airborne, pop-up flights need to be re-scheduled, thereby also impacting other aircraft. The results therefore indicate that, even when the take-off time accuracy is two minutes, it is better to not pre-plan pop-up flights. This effect worsened for larger take-off time inaccuracies.

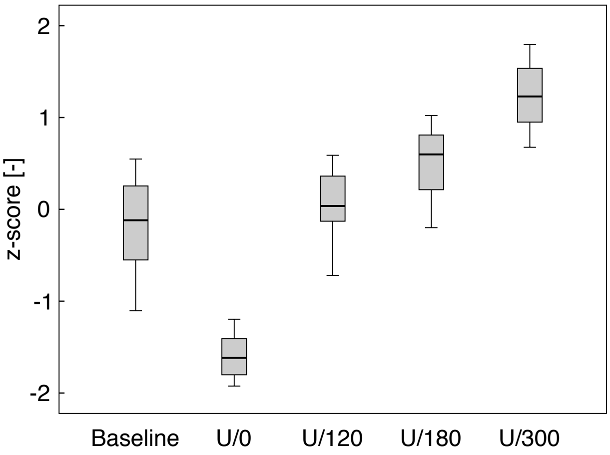

Fig. 10. STA revisions (U=Uncertainty Departure Time Estimate).

3) *Flight Crew and ATC Task Load*

The results in terms of STA revisions are shown in Fig. 10. In this figure, the result of the baseline situation (no pre-planning of pop-up flights) is shown along with the results of the four conditions in which pre-planning of pop-up traffic takes place: U/0 (pop-up aircraft departs exactly at its estimated time), U/120 (departs 2 minutes later), U/180 (departs 3 minutes later) and U/300 (departs 5 minutes later). The main effects tests revealed a significant effect of pre-planning on STA revisions ($\chi^2(4) = 41.24, p < 0.01$). Post-hoc tests revealed a significant improvement between the perfect pre-planning condition (U/0) and the baseline ($p < 0.01$), with an average improvement of 38% (see Table IV).

Significant differences were also found between the perfect pre-planning condition and deteriorated condition U/120 ($z = 3.06, p < 0.01$), as well as between the baseline condition and the deteriorated precision conditions for condition U/300 ($z = 3.06, p < 0.01$). In each of these cases, performance worsened with increasing planning uncertainty.

Because with the pre-planning scheduler, pop-up flights have a reserved place in the sequence, schedule revisions are largely unnecessary when flights depart at the estimated time. When take-off time estimate errors are introduced, however, the number of required STA revisions and the number of affected aircraft increase. In this case, pop-up aircraft are pre-planned using the wrong take-off time estimates. Once airborne, they will have to be re-scheduled, possibly also impacting other airborne aircraft. The larger the estimate errors, the more revisions and impacted aircraft, as can be seen in Fig. 10.

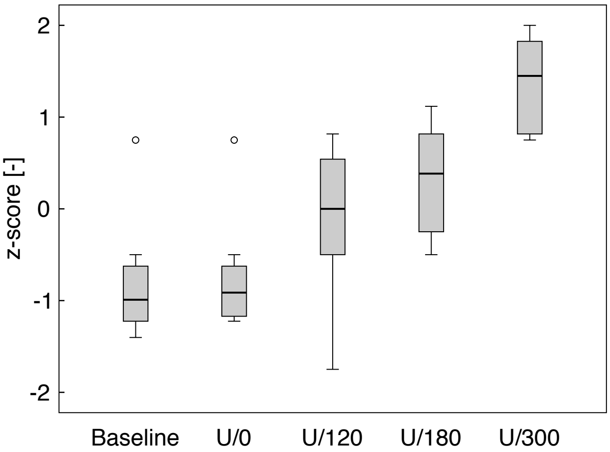

Fig. 11. Disturbed descents (U=Uncertainty Departure Time Estimate).

Fig. 11 shows the number of disturbed descents. In this figure, the result of the baseline situation (no pre-planning of pop-up flights) is shown along with the results of the four conditions in which pre-planning of pop-up traffic takes place: U/0 (pop-up aircraft departs exactly at its estimated time), U/120 (departs 2 minutes later), U/180 (departs 3 minutes later) and U/300 (departs 5 minutes later). Here, a main effects test revealed a significant impact of pre-planning on the number of disturbed descents ($\chi^2(4) = 31.96, p < 0.01$). Post-hoc tests, however, only revealed significant differences between the baseline and U/300 conditions ($z = 2,95, p < 0.01$). Nevertheless, it can be seen in Fig. 11 that when the take-off time estimate errors increase, there is a tendency towards a growing number of disturbed descents. The actual effect is small because most pop-up aircraft depart further away from the airport than the Top of Descent (ToD) of airborne flights, as can be seen in Table IV.

V. DISCUSSION

The experiments focused on assessing the effect of pop-up flights on the (extended) arrival management process. In addition, it was analysed whether this negative impact could be mitigated by pre-planning pop-up flights prior to departure. Statistical results, both from the main and post-hoc tests, are summarized in Table V for both experiments.

A. Effect of Pop-Up Flights

For E-AMAN, several trends and tendencies can be observed. First, when the occurrence of pop-up flights increases, there is a large negative and significant effect on flight crew and ATC task load, as well as on sequence stability. This is revealed by the increased number of STA revisions and arrival sequence position changes. When pop-up occurrence increases, the number of disturbed descents does too, although this is not a statistically significant result. Moreover, the corresponding effect magnitude is considered negligible. The average low-level delay absorption is not significantly affected by pop-up scaling. Delay energy cost grows when pop-up occurrence increases. In case the pop-up occurrence were doubled, this cost increase would be significant and largely negative. As hypothesized, the occurrence of pop-up flights has a significant and large negative effect on delay (cost) (H1-1) and sequence stability (H1-2), as well as on flight crew and ATC task load (H1-3). These effects and tendencies are clear when observing current levels of pop-up occurrence within the context of E-AMAN. In addition, these issues grow when the pop-up occurrence is doubled, clearly illustrating that pop-up flights negatively affect the extended arrival management process.

Most of these effects are observed in the context of AMAN as well. However, due to the lower occurrence of pop-up flights, the impact is smaller when compared to E-AMAN, and is therefore often not statistically significant. This finding confirms hypothesis H1-4.

B. Horizon Extension

It was also assessed whether a horizon extension, from the AMAN to E-AMAN context, is beneficial in terms of overall system performance. On the one hand, this extension positively affects the delay (cost): the required low-level delay absorption is reduced by 17%. Also the delay energy cost reduced by 22% on average. The number of disturbed descents was reduced to nearly zero, although their occurrence (9%) in AMAN is not considered problematic either. On the other hand, the number of STA revisions increases by a factor 27 when extending the horizon. In addition, the number of position changes is negatively impacted, on average by a factor 6. Obviously these effects are induced by the increased occurrence of pop-up flights in E-AMAN. As hypothesized, the horizon extension has a clear benefit in terms of delay (cost) (H1-5), whereas sequence stability (H1-6), as well as flight crew and ATC task load (H1-7), are negatively affected.

C. Alternative Scheduler

The advantages of an AMAN horizon extension are large, and therefore should be pursued. However, mitigation actions need to be taken to limit the observed negative effects of increased pop-up occurrence. It was analysed whether pre-planning pop-up flights prior to departure, using their take-off time estimates, is beneficial. An alternative scheduler was developed that explicitly schedules and pre-plans pop-up flights prior to departure. By comparing this alternative pre-planning scheduler with the baseline scheduler, it could be assessed whether pre-planning pop-up flights is beneficial. It was observed that pre-planning is beneficial, but only when there are no take-off time estimate inaccuracies. In this case, the number of STA revisions could be reduced by 38%. This is positive in terms of flight crew and ATC task load. In addition, both schedulers result in similar performance in terms of average low-level delay absorption, the number of position changes, disturbed descents and the delay energy cost. Pre-planning is therefore mainly beneficial in improving task load, and thereby outperforms the baseline scheduler, as hypothesized (H2-1). It is however important to realize that it relies on perfectly accurate and reliable take-off time estimates. In reality, however, flights are often delayed prior to departure - in the order of minutes - for various reasons, and therefore this requirement seems unrealistic.

If the take-off time estimate error increases to 120 seconds, the conclusions change. In terms of the number of STA revisions, the alternative scheduler no longer outperforms the baseline scheduler. In addition, the scheduler performs statistically significantly worse in terms of position changes (+50%).

TABLE V
EXPERIMENT STATISTICAL RESULTS

	Experiment I			Experiment II					
	Main test	Post-hoc tests		Main test	Post-hoc tests				
		E/0-E/100	E/100-E/200		B-U/0	U/0-U/120	B-U/120	B-U/180	B-U/300
Average low-level delay abs.	-			+	-	+	-	-	-
Delay energy cost	+	-	+	+	-	+	-	-	-
Position changes	+	+	+	+	-	+	+	+	+
STA revisions	+	+	+	+	+	+	-	-	+
Disturbed descents	+	-	-	+	-	-	-	-	+

+ significant - not significant

Moreover, there is a tendency - although not statistically significant - which indicates an increase in the average required low-level delay absorption (+5%) and the delay energy cost (+2%). Even when estimate accuracies of 2 minutes would be achievable, the results show that it is better to discard the information and not pre-plan pop-up flights prior to take-off. Hypothesis H2-2 is therefore not supported by the results. On overall, the alternative scheduler's performance deteriorates with increasing take-off time estimate errors. With an estimate uncertainty of five minutes, the scheduler is outperformed by the baseline scheduler in all observed metrics.

These accuracy requirements and conclusions depend on the scheduler being used by the arrival manager, and could possibly change when the scheduling algorithm is modified. In both schedulers used during the experiments, a schedule revision was triggered once a pop-up flight actually departed. This revision could impact airborne flights, as is the case in most European AMAN systems nowadays. One could however consider adapting the scheduler algorithms, such that schedule revisions are not initiated by default once a pop-up flight departs. Rather, if the pop-up flight departs shortly after its scheduled time, there is no schedule revision and it is up to the pop-up flight to arrive at its destination on time anyway.

The research findings are similar to the outcomes of a NASA study [18] on the *Multi-Center Traffic Management Advisor*, the US equivalent of E-AMAN [18]–[22]. In this study, it was examined whether it is beneficial to pre-plan pop-up flights prior to departure. During the study, it was concluded that it is better to discard the inaccurate take-off time estimates for pre-planning purposes. Rather, it is better to schedule the pop-up aircraft only once airborne. [18] These conclusions are consistent to the findings in this research project: discard inaccurate estimates in the context of arrival management, as it only disturbs the process to a larger extent. Accurate estimates are required and can result in overall benefits for AMAN and E-AMAN. However this is only the case when the take-off time estimate errors are actually eliminated. Similar conclusions were found for a study [17] that assessed whether take-off times could be tuned in order to avoid en-route conflicts. As in the context of arrival management, the required accuracies are highly demanding. It was observed that the (positive) effect diluted as the pre-departure estimate uncertainties were increased [17]. Similar to the context of arrival management, very accurate information is required for pre-planning aircraft prior to departure. If this is not the case, it is better to discard the information and not pre-plan pop-up flights.

The Collaborative Decision Making (CDM) and Advanced ATC Tower concepts have proven their value, as take-off time estimates are becoming more reliable and accurate. In the context of Air Traffic Flow & Capacity Management (ATFCM), these improved estimates are crucial for improving predictions on ATC sector counts. Due to the nature of the arrival management process, the required accuracies are substantially higher when compared to ATFCM. The current estimated take-off time window is still in the order of minutes, which implies that the information is insufficiently accurate for using it effectively in AMAN and E-AMAN. When considering that currently take-off time estimate accuracies in the order of five minutes are achievable, it has been shown that it is better to discard the estimates and not pre-plan pop-up flights prior to departure, when the current scheduler algorithms are being used.

VI. CONCLUSION

In BlueSky, an open-source ATM simulator, experiments were performed using an AMAN research model. The effect of pop-up flights, both in the context of arrival management (AMAN) and extended AMAN (E-AMAN), has been analysed for Amsterdam Schiphol Airport. In addition, it was assessed whether it is beneficial to pre-plan pop-up flights, prior to departure, using their take-off time estimates.

Results show that pop-up flights negatively affect the (extended) arrival management process, in terms of flight crew and air traffic control (ATC) task load, sequence stability and delay (cost). When extending the AMAN horizon, the occurrence and effect of pop-up flights grows substantially, such that mitigation actions are needed. Pre-planning pop-up flights is beneficial, mainly in terms of flight crew and ATC task load, but only if the pre-departure take-off time estimates (for pop-up flights) are sufficiently accurate. If the accuracy deteriorates to 2 minutes or more, it is not considered better to pre-plan pop-up flights. On the contrary, when pre-planning using currently achievable estimate accuracies of approximately 5 minutes, it was observed to result in worse overall performance when compared to the situation in which pop-up flights are only considered once airborne. The more deteriorated the estimate accuracies, the larger the negative effects of pre-planning pop-up flights.

These conclusions are applicable for those AMAN schedulers that trigger a schedule revision when pop-up flights depart, as is the case for most European AMANs currently. More research on alternative AMAN scheduler algorithms is required, for incorporating take-off time uncertainties of pop-up flights in an efficient way, and for dealing with pop-up flights efficiently in the context of E-AMAN.

ACKNOWLEDGMENTS

The authors would like to thank To70, an international aviation consultancy organization located in the Netherlands, for giving Alexander the opportunity to present his work at the ATM Seminar.

REFERENCES

[1] G.H. Visser, "Terminal area traffic management," *Progress in Aerospace Sciences*, vol. 28, no. 4, pp. 323–368, 1991.

[2] N. Hasevoets and P. Conroy, "Aman status review 2010," Eurocontrol, Tech. Rep., 2010.

[3] ——, "Arrival manager: Implementation guidelines and lessons learned," Eurocontrol, Tech. Rep., 2010.

[4] "European atm master plan - 2015," SESAR, Tech. Rep., 2015.

[5] "Aman information extension to en-route sectors: Concept of operations," Eurocontrol, Tech. Rep., 2009.

[6] M. Tielrooij, C. Borst, M. Mulder and D. Nieuwenhuisen, "Supporting arrival management decisions by visualising uncertainty," in *SIDs 2013 - Proceedings of the SESAR Innovation Days*, 2013. [Online]. Available: www.scopus.com

[7] "Ec information annex n: Fabec airspace design," FABEC, Tech. Rep., 2012.

[8] D. Schleicher, E. Jones, D. Dow and R. Coppenbarger, "Improved lateral trajectory prediction through en route air-ground data exchange," in *Proceedings of AIAA's Aircraft Technology, Integration and Operations Technical*, 2002.

[9] I. Lymperopoulos and J. Lygeros, "Improved multi-aircraft ground trajectory prediction for air traffic control," *Journal of Guidance, Control and Dynamics*, vol. 33, no. 2, 2010.

[10] S. Hong and K. Lee, "Trajectory prediction for vectored area navigation arrivals," *Journal of Aerospace Information Systems*, vol. 12, no. 7, 2015.

[11] C. Gong and W. Chan, "Using flight manual data to derive aero-propulsive models for predicting aircraft trajectories," in *Proceedings of AIAA's Aircraft Technology, Integration and Operations Technical*, 2002.

[12] W. Liu and I. Hwang, "Probabilistic trajectory prediction and conflict detection for air traffic control," *Journal of Guidance, Control and Dynamics*, vol. 34, no. 6, 2011.

[13] M. Tielrooij, C. Borst, M.M. van Paassen and M. Mulder, "Predicting arrival time uncertainty from actual flight information," in *Proceedings of 11th USA/Europe Air Traffic Management R&D Seminar*. Delft University of Technology, 2015.

[14] G. Tobaruela, P. Fransen, W. Schuster, W.Y. Ochieng and A. Majumdar, "Air traffic predictability framework - development, performance evaluation and application," *Journal of Air Transport Management*, vol. 39, pp. 48–58, 2014. [Online]. Available: www.scopus.com

[15] J.M. Hoekstra and J. Ellerbroek, "Bluesky atc simulator project: an open-data and open-source approach," in *Proceedings of the 7th International Conference on Research in Air Transportation*, Philadelphia, 2016.

[16] R. Barragán, J. A. Pérez Castán et al., "Optimised algorithms for extended arrival managers," in *Proceedings of 29th Congress of the International Council of the Aeronautical Sciences*. Technical University of Madrid (UPM), 2014.

[17] N. Barnier and C. Allignol, "Trajectory deconfliction with constraint programming," *The Knowledge Engineering Review*, vol. 27, no. 3, 2012.

[18] J. Thipphavong and S. Landry, "The effects of the uncertainty of departures on multi-center traffic management advisor scheduling," in *Proceedings of AIAA 5th Aviation Technology, Integration and Operations Conference*. Arlington: NASA, 2005.

[19] W. Nedell and H. Erzberger, "The traffic management advisor," in *Proceedings of 1990 American Control Conference*. NASA, 1990.

[20] R. Coppenbarger, R. Lanier, D. Sweet and S. Dorsky, "Design and development of en route descent advisor (eda) for conflict-free arrival metering," in *Proceedings of AIAA Guidance, Navigation and Control Conference*. Providence: NASA/FAA, 2004.

[21] R. Slattery and Y. Zhao, "Trajectory synthesis for air traffic automation," *Journal of Guidance, Control and Dynamics*, vol. 20, no. 2, 1997.

[22] M. Jackson, Y. Zhao and R. Slattery, "Sensitivity of trajectory prediction in air traffic management," *Journal of Guidance, Control and Dynamics*, vol. 22, no. 2, 1999.

AUTHOR BIOGRAPHIES

Alexander Vanwelsenaere received his MSc in Aerospace Engineering from Delft University of Technology in September 2016. His graduation work focused on assessing the effect of pop-up flights on the extended arrival manager, a collaboration project between Delft University and Air Traffic Control the Netherlands. In addition, he also concluded traineeships at Eurocontrol, in the departments of safety and air traffic services. He recently joined To70, an aviation consultancy firm located in the Netherlands, where he continues to work on (inter)national aviation-related projects in the field of ATM, airports, safety and the environment.

Joost Ellerbroek received the M.Sc. (2007) and Ph.D. (2013) degrees in aerospace engineering from the Delft University of Technology, The Netherlands, with the Control and Simulation section, where he is currently working as an Assistant Professor. His research interests lie in the field of Air Traffic Management, and include topics such as the analysis of airspace complexity and capacity, data science in ATM, the design and analysis of conflict detection and resolution algorithms, and human-automation interaction.

Jacco Hoekstra holds a chair in CNS/ATM in the Control & Simulation department of the faculty of Aerospace Engineering at the TU Delft. He obtained his MSc, together with his pilot license in 1991, after which he started working in the Handling Qualities & Flight Simulation department at the Dutch National Aerospace Laboratory NLR. He worked on and led many avionics and ATM research projects on topics like Data link, 4D TBO and Airborne Separation Assurance in co-operation with FAA, NASA en European partners. Using the results of this research, in 2001 he obtained his PhD. At NLR he led several departments after which he became head of the Air Transport research division. In 2007 he moved to the TU Delft, initially to serve two terms as dean of the Aerospace Engineering faculty. He was subsequently asked to set up the CNS/ATM chair at this faculty in 2013.

Evert Westerveld holds a master degree in Aerospace Engineering from Delft University of Technology. He started his career as a specialist Stability and Control at Fokker Aircraft Inc. In 1996 he joined LVNL, air traffic control the Netherlands where he became manager ATM research and subsequently manager of the Knowledge & Development Centre. Evert Westerveld has focused his professional career on the development of the arrival management domain in which he has supported and led national and European research and development projects.