Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)

Evaluation of a Dynamic Weather-Avoidance Rerouting Tool in Adjacent-Center Arrival Metering

Miwa Hayashi, Doug Isaacson, and Huabin Tang
NASA Ames Research Center
Moffett Field, CA, USA

***Abstract*—Dynamic Reroutes for Arrivals in Weather (DRAW) is a NASA-developed decision-support tool for Traffic Management Coordinators (TMCs) at the Federal Aviation Administration's Air Route Traffic Control Centers ("*Centers*"). DRAW proposes weather-avoidance reroutes for en route arrival flights subject to metering restrictions when transitioning into a busy terminal airspace. The prior DRAW study demonstrated that TMCs' use of DRAW promotes earlier reroutes of arrivals, and reduces the number of routes conflicting with weather in the Center. The present paper focuses on how DRAW benefits metering delivery accuracy when schedule freeze horizon distance was altered. A human-in-the-loop simulation was conducted at NASA Ames Research Center in October-November 2018, where retired TMCs and controllers performed simulated metering operations for southeast arrivals through the Atlanta and Jacksonville Centers to the Hartsfield-Jackson Atlanta International Airport during convective weather periods. Results demonstrated that DRAW use reduced the frequency of manual adjustments of Scheduled Times of Arrival and lowered TMC workload. DRAW use also made the metering accuracy, the number of reroute amendments after the freeze horizon, and the en route sector controller workload more robust to the effect of different freeze horizon distance.**

I. INTRODUCTION

A. Background

When arrival demand exceeds the capacity of a major airport, time-based arrival metering restrictions are implemented for en route flights entering the terminal airspace. These restrictions reduce large path-stretches and holds in the congested airspace and maintain a manageable flow rate and efficient sequencing at the runway. Successful arrival metering also increases schedule predictability and enables efficient resource planning in the rest of the National Airspace System operations. To support arrival metering operations, Traffic Management Advisor (TMA) is used in the US [1], and Arrival Manager (AMAN) is used in Europe [2-4].

In the US, TMA assigns a scheduled time of arrival (STA) to an individual arrival for the designated *meter fix* (MFx) located just before entering the Terminal Radar Approach Control (TRACON) airspace. STAs are assigned largely on a first-come first-served basis based on the flight's estimated time of arrival (ETA) and the planned runway sequence, as well as other user-specified constraints. When the flight passes the schedule freeze horizon (FH), the STA is frozen.

Inside the FH, each flight's target *delay*, defined as STA – ETA, is presented and continuously updated on the sector controller's screen at the Air Route Traffic Control Center (henceforth, *Center*). A positive target delay means the flight needs to be delayed, and a negative one means it needs to be expedited. The controller issues speed changes and vectors to null the target delay.

Weather can challenge sustaining metering, because any tactical deviations around the weather affect the flight's ETA, and, in turn, could make the STA, already frozen, hard to achieve. When this happens to many flights, and maintaining the metering becomes difficult, the TMC may suspend metering and switch to Miles-in-Trail (MIT) restrictions, which would reduce runway throughput [5].

B. Dynamic Routes for Arrivals in Weather (DRAW)

DRAW is a NASA-developed decision-support tool for Center TMCs to aid the arrival metering operations in severe weather [6]. DRAW provides the TMC with arrival reroute advisories for 1) weather avoidance and 2) increased efficiency, such as shortcuts through an area previously forecasted for weather that becomes available, or reroutes through an alternate MFx. By assisting the TMC in implementing more of such reroutes earlier, DRAW aims to reduce the need for last-minute tactical maneuvers near the MFx and keep the frozen STAs more feasible, in order to eventually enable metering under a broader range of weather conditions.

The DRAW software resides in NASA's Center/TRACON Automation System (CTAS) tool suite. It leverages various trajectory optimization capabilities available in the CTAS, such as the research version of TMA for scheduling arrivals, as well as its weather database, such as the Corridor Integrated Weather System (CIWS) and Convective Weather Avoidance Model (CWAM). DRAW uses the CIWS to depict the current weather on the CTAS's planview graphical user interface (PGUI) display. It uses CWAM Weather Avoidance Field (WAF) polygons for computing weather-avoidance rerouting. CWAM WAFs use the CIWS forecast data as input to calculate the contours of the areas where a certain percentage of pilots are predicted to avoid entering [7]. Using CWAM instead of CIWS helps DRAW to identify reroutes that the pilot and controller are more likely to accept. DRAW currently uses the CWAM WAFs for terminal airspace regions without an altitude cap, so DRAW

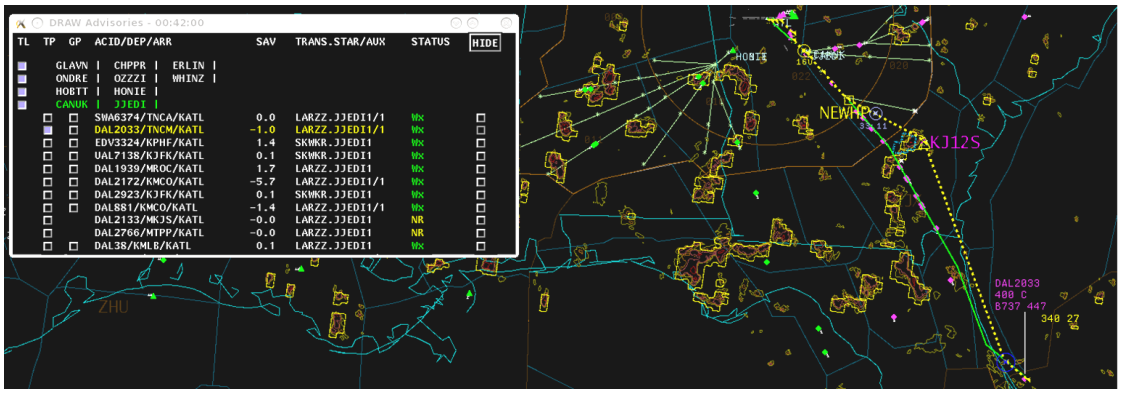

Figure 1. CTAS PGUI display. The DRAW Advisory List window is presented in the upper left corner. The green solid lines show the current Flight Plan route, and the yellow dotted lines the DRAW reroute advisory. CIWS weather contours are shown.

routes do not fly over the WAFs [8]. In addition to weather cells, custom route-avoidance polygons can be drawn on the PGUI display to prohibit DRAW from rerouting through them.

Every 12-seconds (i.e., after each radar track update), DRAW performs the following two-step computation process:

1. Search for a reroute that saves at least 5 minutes (user defined) of wind-corrected flight time without causing any weather conflict. Candidates include routes going through an alternate MFx. If a time-saving reroute is found, the search ends. If not, it proceeds to Step 2.

2. If the Flight Plan route has a weather conflict, find a weather-avoidance reroute. Auxiliary waypoints are inserted as needed to deviate around the CWAM WAFs.

The resulting DRAW reroute advisories are posted in the DRAW Advisory List on the PGUI display, showing the flight call signs, predicted time savings, Standard Terminal Arrival Route (STAR) transitions, auxiliary waypoints (if any), and advisory status information (e.g., weather avoidance, alternate MFx, alternate MFx for weather, or no weather-avoidance resolution found) (Fig. 1, upper-left corner). The TMC can select any of the advisories in the DRAW Advisory List (or a group of advisories on a similar route, if DRAW identifies any) to graphically evaluate on the PGUI map.

Optionally, the TMC can modify the advised reroute using the CTAS's Trial Planning (TP) graphical user interface. The CTAS's TP interface allows the route to be modified by drag-and-drop, and the route automatically snaps to the nearest waypoint. When DRAW is provided, the TP also displays the conflicting CWAM WAF forecast polygons (if any) as the route is being dragged. Dragging the route also causes DRAW to display and continuously update the ETA, STA, and delay propagations for subsequent flights, which would result if the reroute were accepted, on the CTAS's timeline graphical user interface (TGUI) display (Fig. 2).

Once the TMC finishes the evaluation and decides to accept the reroute, he/she sends the Flight Plan amendment information electronically to the sector controller via the Airborne Rerouting (ABRR) functionality. The sector controller may accept or reject the route amendment. If ABRR is not available, the TMC must work with the Area Supervisor to pass the amendment to the corresponding sector controller.

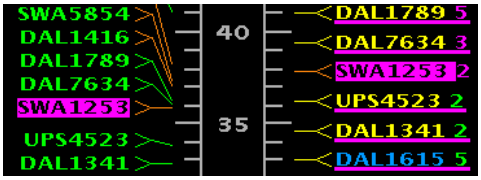

Figure 2. CTAS TGUI display. The numbers in the middle of the vertical timeline are the times to the MFx. ETAs are displayed on the left (with green aircraft ID tags), and STAs on the right (yellow if unfrozen, or blue if frozen). The target delay values are shown on the right side of STAs. The magenta-highlighted tags are the ETA and STA of the DRAW reroute being evaluated, and the magenta numerals on the right are the target delays that would result. The other flights that currently have a DRAW advisory also are shown with a magenta underline.

C. *Related Previous Work*

1) Dynamic Route-Planning in Convective Weather: MITRE Corp. has developed Advanced Flight-Specific Trajectories (AFST), a tool similar to DRAW, which proposes to a TMC user a weather-avoidance route for en route arrivals to help sustain metering during severe weather periods [9]. Unlike DRAW, AFST offers reroutes only prior to the FH, and the weather-avoidance reroutes are formulated based on a historical-route database. They conducted a human-in-the-loop exercise in 2017 with retired TMC participants using Dallas-Fort Worth International airport (KDFW) arrival traffic data. They estimated that using AFST would save US$1.23 million annually for KDFW arrival operations alone.

NASA's Dynamic Weather Routes (DWR), DRAW's predecessor, proposes to airline dispatchers more efficient routes around weather for flights in their en route phase. DWR was field-tested at the American Airlines Integrated Operations Center in Dallas-Fort Worth, Texas, in 2012 and 2013-2014 [10-11]. NASA's Multi-Flight Common Route (MFCR) is another variant of DWR, adapted for Center TMCs' use [12]. MFCR detects a group of en route flights that can be rerouted together for a shortcut through a weather-free area previously forecast to be occupied by convective weather. Grouping flights helped reduce controller workload for implementing the reroutes. However, neither DWR nor MFCR offers any assistance for arrival metering.

2) Arrival Metering Delivery Accuracy: The *acceptable* accuracies reported by literature vary. In NASA's Air Traffic

Management Technology Demonstration (ATD-1) human-in-the-loop simulation studies, 30 to 40 seconds was used as the acceptable upper limit of the actual time of arrival (ATA) error at the MFx, assuming that TRACON could absorb 15 to 60 seconds of delay using speed control only [13-15]. Sharma reported that, in their Operational Integration Assessment simulation study of NASA's Terminal Sequencing and Spacing system, 88% of flights attained an ATA error within ±30 seconds, and the standard deviation was 24 seconds using conventional metering techniques (or 89% and 26 seconds, respectively, with their Ground-based Interval Management–Spacing speed advisories) [15].

Based on their numerical simulation studies, both Vandevenne [16] and Thipphavong [17] concluded that the maximum acceptable standard deviation of the ATA error for ensuring the metering sequence and proper interval management was half of the delay margin (i.e., the maximum absorbable delay) available in the TRACON area. Shresta derived that the maximum standard deviation of the ATA error at the MFx acceptable for the Atlanta TRACON was 60 seconds if full vectoring and speed control were allowed, or 10 seconds if only speed control was used [18]. In the present study, the accuracy performances suggested by [15] and [18] are referred to in the Result section.

3) Previous DRAW Evaluation: An earlier human-in-the-loop simulation evaluation study of DRAW was conducted at NASA Ames Research Center in 2016 using the KDFW arrival traffic data through the western half of the Fort Worth Center (ZFW) airspace [19]. The study found that the TMCs rerouted flights about 16 minutes earlier when using DRAW compared to not using it. Also, use of DRAW reduced the number of flights that still had residual weather conflicts in the ZFW airspace when crossing the ZFW Center boundary. Despite these positive indications of DRAW's successful operation, the study did not find any evidence that DRAW improved arrival metering performance.

D. Objectives and Scope

The primary objective of the current study was to investigate how arrival metering performance in weather was affected by:

1. Use of DRAW
2. Interaction of use of DRAW and the FH distance

More specifically, the following performance measures in arrival performance were examined: 1) ATA errors, 2) the frequency of Flight Plan route amendments issued before and after the FH, 3) the frequency of manual STA adjustments, 4) controller workload, and 5) TMC workload.

The first bullet above (i.e., DRAW effect) intended to fill the gap in the previous DRAW study [19]. The second bullet (i.e., DRAW × FH interaction effect) was a new topic. A FH located far from the MFx provides greater delay-absorption capacity due to the long distance between the FH and the MFx, whereas a FH located close to the MFx yields a more accurate ETA due to the shorter distance to the MFx and fewer uncertainties (e.g., departures from inside the FH, or *pop-ups* [20]). A hypothesis was that using DRAW would make the arrival metering performance more robust to different FH distances than not using it.

A secondary objective of the study was to observe any inter-Center coordination required for weather avoidance and arrival metering operations. Unlike the previous DRAW study, the current study staffed the adjacent Center positions to allow simulating and reasonably comparing the effects of the different FH distances, both located in the adjacent Center.

In order to serve the above objectives efficiently, the study focused only on a single arrival gate of the Hartsfield-Jackson Atlanta International Airport (KATL). KATL arrival traffic to the other gates was also present to pressure the runway schedule but was not actively attended by the study participants. Unlike real operation, no manual offloading of arrival traffic from one MFx to another was permitted. Hence, DRAW's capability to reroute via an alternate MFx was disabled in this study. The reduced scope ensured the same set of arrivals went to the MFx of interest (*JJEDI*) in every run and allowed for fair comparison between different run conditions.

II. METHODS

A. Airspace

The simulated airspace is illustrated in Fig. 3. All southeast arrivals were routed to the JJEDI MFx. Four high sectors in ZJX and ZTL (ZJX58, ZJX50, ZJX66, and ZTL20) and one low sector in ZTL (ZTL19) were staffed with the controller and pilot participants. (To reduce the number of required participants, the ZJX58 sector in the simulation combined two high sectors, ZJX75 and ZJX58. In addition, the three ZJX high sectors included the ultra-high sectors above them covering FL330 and above. The right-hand section of ZJX58 in Fig. 3, protruding over the ocean, is the ultra-high sector only.) A ghost-sector controller handled sector handoffs to/from outside these five sectors. All flights inside the Atlanta TRACON flew automatically along the approach path to Runway 28. Data in the TRACON were not analyzed in this study.

The two FH distances tested were 275 nmi and 175 nmi in radius from the JJEDI MFx , denoted as *Far* and *Near* FH, respectively. Only one of the distances was used in each run. The

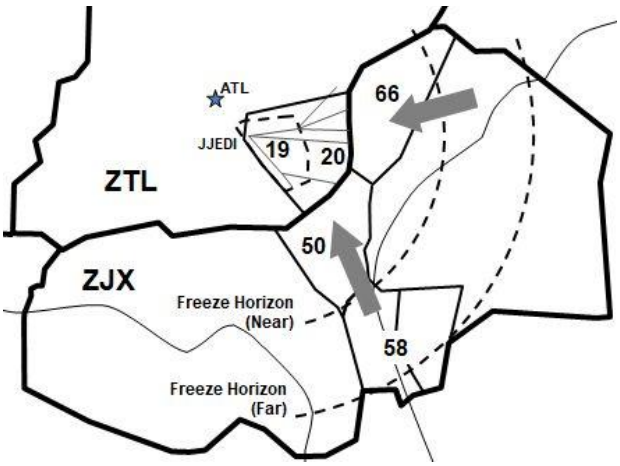

Figure 3. ZTL and ZJX airspace simulated. The thick gray arrows depict traffic flows. The thin curved lines show the US coastlines.

Far FH is similar to what is being used in the present-day KATL metering operation. Note that when the Far FH was used, all five sectors participated in the metering, whereas when the Near FH was used, all but ZJX58 participated. Non-participation means that the STAs were not yet frozen, and the sector did not delay or expedite flights for the metering purpose.

B. Participants

Two retired ZTL TMCs and two retired ZJX TMCs participated as the TMCs in the two-week study. A pair of one ZTL TMC and one ZJX TMC worked during the first week (Team A), and the other ZTL TMC and ZJX TMC worked during the second week (Team B). Their Center TMC experience was 11.8 years on average, ranging from 5 to 14 years. They retired within 4 years prior to the experiment dates on average, ranging from 0 to 9 years.

Seven retired Center sector controllers participated in the study, of which two were retired ZTL controllers, one was a retired ZJX controller, and the remaining four were retired Oakland Center (ZOA) controllers. One of the two ZTL controllers participated in the first week (Team A), and the other in the second week (Team B). The ZTL controllers always staffed the ZTL19 low sector, the most critical sector for the metering operation. The remaining ZJX and ZOA controllers participated in both weeks (i.e., both Teams A and B), each staying at the same sector throughout the two weeks. The retired ZJX sector controller controlled the ZJX50 high sector. One ZOA controller performed the ghost-sector controller role. These seven controllers' Center experience was 28.9 years on average, ranging from 25 to 33 years. All retired from the Center within 5 years on average, ranging from 0 to 12 years. Six pilot participants were recruited from the local general-aviation community to perform pseudo-pilot duties in the simulation.

C. Laboratory Setup

The study was conducted in the ATC Laboratory at NASA Ames Research Center. The lab was configured to provide one TMC station, six sector-controller stations, six pilot stations, and one tactical weather-avoidance station.

The TMC station included the PGUI and TGUI displays generated by CTAS (Fig. 4). Time-lapse weather-radar images of the last 120 minutes were played on a loop on a laptop-computer monitor.

Figure 4. TMC station. PGUI on the left monitor, and TGUI on the right. Weather radar images are animated on the laptop-computer monitor.

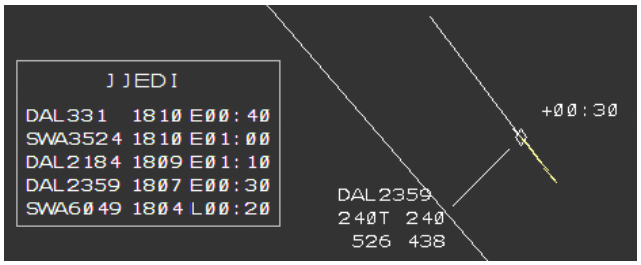

Figure 5. ERAM Meter List and Delay Countdown Timer ("+00:30"). "E" in the Meter List means early, whereas "L" means late.

The sector-controller stations offered En Route Automation Modernization (ERAM) displays emulated by NASA's Multi-Aircraft Control System (MACS). On the ERAM display of the sectors participating in metering, a Meter List with the flights' STAs and Delay Countdown Timer (DCT), as well as a DCT next to the flight's symbol, were presented (Fig. 5). To improve metering accuracy, DCTs were shown in the tens-of-seconds format ("mm:ss") rather than the rounded-minute or truncated-minute format ("mm") [14]. An "E" in the Meter List DCT or "+" in the DCT next to the flight symbol indicates that the flight is early, whereas an "L" or "-" indicates that the flight is late.

The pilot stations showed the MACS pseudo-pilot user interface, which let them execute the controller instructions for a flight in their assigned sector. The controllers and pilots wore headsets and talked over simulated radio.

In addition, the tactical weather avoidance position was set up next to the pilot stations, staffed by two confederate members. They monitored the PGUI display for any flights heading toward a CIWS weather cell and instructed the pilot who owned the flight to request a weather-avoidance deviation from the controller. This helped the pilot, who was in charge of multiple flights, to request such deviations in a timely manner, as well as to make the timing of the requests consistent among the sectors and different run conditions, normally 80 to 100 nmi prior to the weather conflict.

D. Traffic and Weather

The total length of each run was 2 hours. A single traffic scenario was used for all of the runs, with the aircraft call signs shuffled among runs. NASA's Air Traffic Management–eXploration (ATM-X) Testbed capability [21] was used to extract the base traffic scenario under clear-weather and high-traffic-volume conditions recorded on December 23, 2017. Non-arrivals (i.e., departures and overflights) were removed to make the scenario arrivals-only. Then, traffic density was increased by shifting the two peak times closer and adding more flights merging from the north or northeast. For the TMA scheduler, the airport arrival rate was set at 130 aircraft per hour.

To add variations, two weather scenarios providing weather of roughly comparable severity along the JJEDI arrival paths were used. These scenarios were generated based on the actual and forecast weather data recorded on June 30, 2018 (Scenario W1) and May 17, 2018 (W2). (The weather cells in W2 were shifted to the east by about 750 miles to place them over these arrival paths.)

E. Experiment Design

Sixteen test-matrix runs and two baseline runs were conducted in a two-week period. The main test-matrix design was 2×2×2×2, including:

- DRAW Condition (DRAW vs. No-DRAW)
- FH (Far vs. Near)
- Weather (W1 vs. W2)
- Team (Team A vs. B)

Table I lists what functions were available in DRAW and No-DRAW runs. None of the functions in Table I is available in the current ERAM monitors used at Centers. Thus, the No-DRAW runs were not meant to represent the present-day condition. Note also that the function differences were relevant only to the TMCs; i.e., the controllers had no way to know whether the TMCs were using DRAW or not in each run.

Table I. DRAW conditions

Function	DRAW	No-DRAW
DRAW advisory	X	
Current CIWS weather on PGUI	X	X
TP: Drag and drop to reroute	X	X
TP: Forecast CWAM conflicts	X	
TP: Metering impact on TGUI	X	

TP = Trial Planning

Eight test-matrix runs were performed each week. The order of DRAW Conditions was counterbalanced both between and within Teams, whereas the orders of FH and Weather were counterbalanced only between Teams, not within, to avoid repeating the same weather or FH condition on the same day.

The baseline runs had no weather, so the TMCs provided no reroutes. The purpose of the baseline runs was to assess the arrival metering performance under clear-weather conditions. One baseline was run in the first week with Far FH and Team A, and the other was in the second week with Near FH and Team B. Both baseline runs were conducted in the middle of each week.

F. Procedure

A classroom briefing was provided to the participants on Monday morning of each week, followed by two hands-on training runs in the laboratory. The first run was conducted later on Monday. On the following days, two runs were performed per day: one in the morning and one in the afternoon. At the end of the Friday afternoon run, a debriefing was held to obtain verbal comments from the participants.

Before each run started, a DRAW researcher gave the ZTL and ZJX TMCs a weather briefing about the convective forecast and showed them a time-lapse playback from one-hour of CIWS reflectivity. Then, the TMCs performed the run from 0 through 110 scenario elapsed minutes. The controllers and pilots participated from 15 through 120 scenario elapsed minutes, because no aircraft were in the sectors at the beginning of the scenario.

Only one TMC station was provided for the two TMCs. The ZTL TMC was the designated user of this station and consulted the ZJX TMC for weather-avoidance reroute planning within the ZJX airspace. Each TMC was allowed to walk to the controllers of their own Center to negotiate and/or coordinate reroutes. (In the field, the TMC would typically talk to the Area Supervisor, not directly to the controller, but this process was abbreviated in this simulation.) Once the TMCs decided to implement a reroute, since ABRR functionality was not available in this simulation, the ZJX TMC talked to the corresponding ZJX controller, and the controller manually amended the Flight Plan route via the ERAM command.

The controllers in metering sectors were instructed to reduce the DCTs to the sector target value before handing off the flight to the next sector. The sector target values were 6 minutes for ZJX58, 4 minutes for ZJX50 and ZJX66, 2 minutes for ZTL20, and 0 minutes for ZTL19. (The target value was not 0 for all the sectors, due to how the DCTs were displayed in this simulation.) If a flight had a DCT less than the target delay, the sectors other than ZTL19 were instructed not to expedite the flight for metering purposes.

There were three methods to manually adjust a frozen STA:

1. Controller *swaps* two flights' STAs via an ERAM command
2. TMC adjusts an STA along a TGUI timeline
3. TMC reschedules the Meter List (*ripples the list*) via a TGUI command

The TMCs and controllers were allowed to manually adjust the STAs as much as needed at their discretion: the frequencies of these adjustments were a part of the measurements. The only restriction was that, if a flight was already owned by a non-ghost sector, TMC adjustment of an STA on a TGUI (the second method above) was allowed only at the request of the controller.

Starting from 10 minutes into the scenario, the TMC station's PGUI and TGUI monitors were covered with a large foam board for 5 minutes, then opened for 5 minutes. This *blackout* cycle was repeated until 110 scenario elapsed minutes. The TMC-monitor blackout periods were meant to artificially increase the time pressure while they worked on only one arrival corner post in this simulation, instead of four as in the field. During each blackout period, the TMCs responded to a questionnaire on a tablet. The ZTL TMCs' questionnaire collected their self-assessed real-time workload rating on the Subjective Workload Assessment Technique (SWAT) scale [22]. The ZJX TMCs' workload were not collected because they were not the DRAW users. The controllers reported their workload ratings on the NASA Task Load Index (TLX) scales on a tablet questionnaire at the end of each run [23].

III. Results

A. ATA Error

1) ATA Error with Respect to STA at FH (E_{fh}): This error is defined as $E_{fh} = STA_{fh} - ATA$, where STA_{fh} is the STA assigned to the flight at the time of the FH crossing, and ATA is the actual time of arrival at the MFx. The magnitude of this error is what DRAW directly tries to minimize (when the route amendment could be issued before the FH). In all analyses below, the absolute values of the error, $|E_{fh}|$, were used. $|E_{fh}|$ is not the final metering performance (see the next section). However, a smaller $|E_{fh}|$ is still desirable, because the STA_{fh} assigned by the TMA

scheduler is coordinated with the other corner posts' arrivals, as well as optimized for interval management and runway sequencing. Reducing this error also improves system-wide schedule predictability. Higher E_{fh} suggests more interventions are required by the en route sector controllers and TMCs.

To assess effects of the experiment parameters, the arrival flights were grouped into three 30-minute segments by their ATA—30 to 60, 60 to 90, and 90 to 120 scenario-elapsed minutes—denoted as segment k = 1, 2, and 3, respectively. Then, the mean and standard deviation of $|E_{fh}|$ within each segment k, denoted as $e_{fh,k}$ and $s_{fh,k}$ (k = 1, 2, 3), respectively, were computed. The reasoning was that consecutive arrivals' ATA errors may have dependency on each other (i.e., a large ATA error is likely followed by another large ATA error), but their means over 30-minute segments are likely more independent. The segmentation also helped to compartmentalize impacts from occasional outliers. Each segment included about 19 arrivals.

The segment means and standard deviations, $e_{fh,k}$ and $s_{fh,k}$, of non-baseline runs were then subjected to a linear regression analysis. Table II lists the main and interaction effects included in the model. Segment effect represents the three 30-minute segments defined above. Participant effect was Team A (first week) versus. B (second week). Run Group effect was included to account for any potential learning or fatigue effect within each week. All effects were categorical.

Table II. Regression model

Effect	Description
DRAW	DRAW or No-DRAW
Segment	Segment in each run (See the text)
Weather	Weather 1 or 2
FH	Far or Near FH location
Run Group	1 (1st half of each week) or 2 (2nd half)
Participant	See the text
Weather x FH	Interaction between Weather and FH
DRAW x Segment	Interaction between DRAW and Segment
DRAW x Weather	Interaction between DRAW and Weather
DRAW x FH	Interaction between DRAW and FH
DRAW x Participant	Interaction between DRAW and Participant

For the segment means, $e_{fh,k}$, the DRAW × FH interaction effect was found to be significant (p = 0.023). To visualize the trends, Fig. 6 shows the means and standard errors of $e_{fh,k}$. The plot shows that in No-DRAW runs, $e_{fh,k}$ increased when the Far FH was used rather than the Near FH, whereas in DRAW runs, there was no such visible trend. The analysis also detected that the $e_{fh,k}$ was smaller when the W2 weather scenario was used than when the W1 scenario was used (p = 0.011) (not shown).

Fig. 7 plots the cumulative percentage of the counts of flights with the absolute error, $|E_{fh}|$, equal to or less than the value on the x-axis. Because the regression analysis informed us that the ATA error means had a different effect between the Far and Near FH distances in No-DRAW runs, the curves were plotted separately for each FH distance. For comparison, the Baseline-run curves are also plotted. The plot shows that the three Near-FH runs (the three dotted curves) were somewhat similar to each other, whereas the three Far-FH runs (the three solid curves) exhibited greater differences. The ATA accuracy in No-DRAW

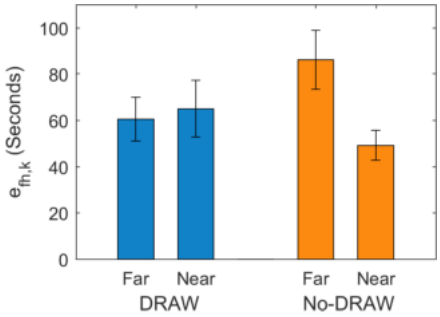

Figure 6. Means and standard errors of ATA-error segment means, $e_{fh,k}$.

Far-FH runs (solid orange curve) was particularly poor compared to the other runs, consistent with the regression finding. Also, the two DRAW run curves (the solid and dotted blue curves) were almost identical, suggesting that DRAW improves robustness for FH distance. The Baseline Near-FH run accuracy was worse than the Baseline Far-FH run. The Baseline curves represent only one run each, so it is hard to conclude what caused the poorer performance—it could be due to the Near FH, as well as the Team B, or just a bad run (e.g., more outliers). Some key readouts of Fig. 7 are summarized in Table III.

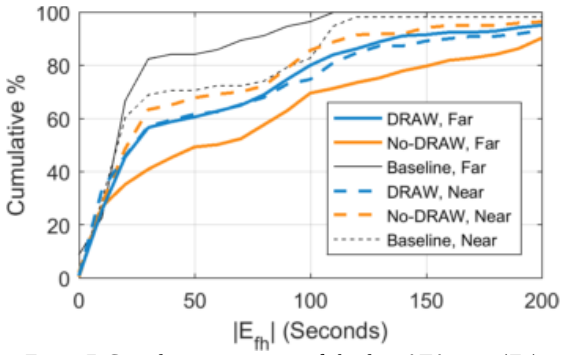

Figure 7. Cumulative percentage of absolute ATA error, $|E_{fh}|$, by DRAW condition and FH location.

Table III. Key readouts of Figure 7

	DRAW		No-DRAW		Baseline	
	Far	Near	Far	Near	Far	Near
30 seconds	57 %	57 %	41 %	63 %	83 %	69 %
40 seconds	59 %	60 %	45 %	65 %	84 %	71 %
90 percentile	144 sec	159 sec	199 sec	114 sec	73 sec	106 sec

For the segment standard deviations, $s_{fh,k}$, regression analysis found that the standard deviations significantly increased in Segment 3 compared to Segment 1 (p = 0.022). No other effect was found significant.

Table IV shows the estimated standard deviations of $|E_{fh}|$ computed using $s_{fh,k}$. The mean of $s_{fh,k}$ is the unbiased estimator of $s_{fh,k}$ (per the Central Limit Theorem) but likely underestimates the population standard deviation due to the small sample size;

Table IV. Estimates of standard deviation of ATA error, $|E_{fh}|$

	Mean of $s_{fh,k}$	Upper bound of 90% CI of $\|E_{fh}\|$ standard deviation (1-tail)
DRAW	68 sec	91 sec
No-DRAW	71 sec	95 sec
Baseline	36 sec	48 sec

i.e., $n \approx 19$. The 90% confidence intervals (CI) were calculated using the critical value of the χ^2 distribution, which also corrects for the effect of the sample size. The upper bounds of the 90% CI, listed in Table IV, are more conservative estimates of the population standard deviations.

Tables III and IV show that both DRAW and No-DRAW runs generally resulted in poor cumulative percentages and large standard deviations of $|E_{fh}|$. That suggests that the controllers and the TMCs had some work to do to attain acceptable metering delivery performance before the MFx.

2) ATA Error with Respect to STA at MFx (E_{mfx}): This error is defined as $E_{mfx} = STA_{mfx} - ATA$, where STA_{mfx} is the STA at the time of the MFx crossing. This is the error that indicates the final metering performance, and the one controllers tried to minimize. Since STA was frozen at the FH, any change between the STA_{fh} and the STA_{mfx} was manually made by the TMC or the controllers.

Analogous to the $|E_{fh}|$ analysis, the segment means and standard deviations of $|E_{mfx}|$, denoted as $e_{mfx,k}$ and $s_{mfx,k}$, respectively, were computed and subjected to the same linear regression analysis. No significant effect was found.

Fig. 8 plots the cumulative percentage of the counts of flights with the absolute ATA error, $|E_{mfx}|$. In this plot, the FH distances were aggregated together. DRAW and No-DRAW runs' ATA accuracies (blue and orange curves) were still not as good as the Baseline runs' (gray curve), but, compared to Fig. 7, they were much improved. The key readouts are listed in Table V. Both DRAW and No-DRAW runs achieved a cumulative percentage for 30 seconds (i.e., 87% and 88%, respectively) that are comparable with the performance reported by [15] without weather (i.e., 88%).

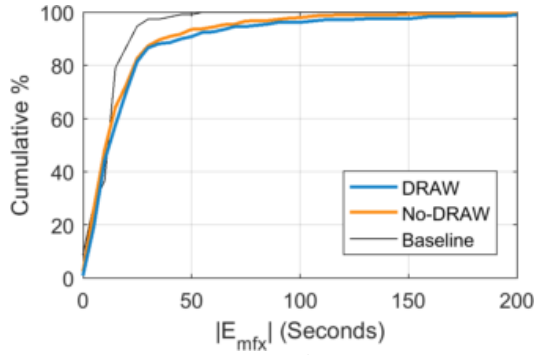

Figure 8. Cumulative percentage of absolute ATA error, $|E_{mfx}|$, by DRAW condition.

Table V. Key readouts of Figure 8

	DRAW	No-DRAW	Baseline
30 seconds	87 %	88 %	97 %
40 seconds	89 %	91 %	98 %
90 percentile	45 sec	37 sec	22 sec

Table VI shows the estimated standard deviations of $|E_{mfx}|$. The DRAW runs (29 seconds) roughly attained the standard deviation comparable to 24 seconds reported by [15], whereas the No-DRAW run did not (42 seconds). Both DRAW and No-DRAW runs met the 60-second criteria suggested by [18]. (Strictly speaking, [15] and [18] did not take the absolute values of the ATA errors. However, the conclusions still appear to roughly hold, as the values in Table VI remain similar even if the computations were repeated without taking the absolute values of the ATA errors.)

Table VI. Estimates of standard deviation of ATA error, $|E_{mfx}|$

	Mean of $s_{mfx,k}$	Upper bound of 90% CI of $\|E_{mfx}\|$ standard deviation (1-tail)
DRAW	22 sec	29 sec
No-DRAW	32 sec	42 sec
Baseline	8.5 sec	11 sec

B. *Flight Plan Route Amendment Locations*

The readers are also reminded that reroutes after the FH could cause a problem for arrival metering accuracy, and thus, utilizing forecast weather information, DRAW aims to reroute arrivals before the FH. To verify this hypothesis, the frequencies of the Flight Plan route amendments before and after the FH in each run were counted. Then, these frequencies were analyzed with the same regression model as shown in Table II, except that Segment and DRAW × Segment effects were omitted. Participant effect represented Team A versus B.

The results found that for the frequencies of the route amendments after the FH, DRAW × FH interaction effect was significant ($p = 0.014$). Fig. 9 shows the means and standard errors of the frequencies of Flight Plan route amendments after the FH per run. The plot shows that, when DRAW was not provided, more route amendments were issued after the FH when the Far FH was used. When DRAW was provided, a similar trend was visible, but not as pronounced. For the frequencies of the route amendments before the FH, no effect was found statistically significant.

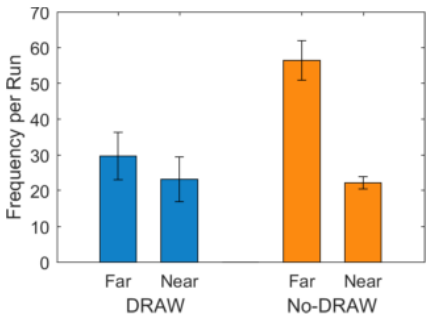

Figure 9. Means and standard errors of frequencies of Flight Plan route amendments after the FH by DRAW × FH

Finding significant effect only in the reroute frequencies *after* the FH implies that the DRAW's forecast CWAM WAFs may have played a major role: when the forecast weather information was not shown, the TMCs often mistakenly rerouted a flight to avoid the *current* weather positions depicted on the screen. These weather positions could be different when the flight arrives there. The problem would worsen when the FH was Far rather than Near simply because the longer distance between the MFx and the FH.

Of course, this benefit of DRAW depends on the accuracy of the CWAM forecast. If the forecast were wrong, DRAW runs may exhibit the same problem as the No-DRAW runs.

C. Manual Adjustments of Frozen STA

The frequencies of 1) STA swaps by the controllers, 2) STA adjustments on TGUI by the TMC, and 3) rippling the list by the TMC per each run were examined. The same linear regression model shown in Table II, except with Segment and DRAW × Segment effects omitted, was applied.

The results showed that the frequency of STA swaps by the controllers was higher in No-DRAW runs than in DRAW runs (p = 0.044, Fig. 10 left). The frequency of STA adjustments on the TGUI by the TMC was higher in No-DRAW runs than in DRAW runs, though this trend was only marginally significant (p = 0.069, Fig. 10 right). The TMC rippled the list only sporadically (in total, five times in the eight DRAW runs, and three times in the eight No-DRAW runs), and no statistically significant effect was found in the frequencies.

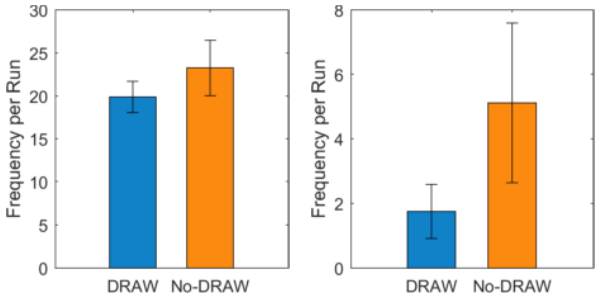

Figure 10. Means and standard errors of STA swaps by controllers (left) and STA adjustments by the TMC (right)

D. Controller Workload

The non-ghost sector controllers' NASA TLX workload ratings were collected on a questionnaire after each run [23]. Each of the six subscale ratings (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) were recorded on an 11-point scale from 0 to 10, where 0 corresponded with the lowest workload level and 10 the highest. Their subscale ratings were simply averaged without weighting to obtain the TLX workload ratings [24].

To account for potentially large between-subject biases, for all the self-reported workload rating analyses in this study, Linear Mixed Model (LMM) regression was used [25]. *R* software (v. 3.5.2) with its *lme4* (v.1.1-19) and *lmerTest* (v. 3.0-1) packages was used for the computation [26-28]. For the controllers' TLX workload ratings, the model in Table II was applied, but Segment and DRAW × Segment effects were omitted. Participant effect represented the five non-ghost sector controllers. Participant and DRAW × Participant interaction effects were treated as *random* effects; all others were treated as *fixed* effects. For the DRAW × Participant interaction effect, a likelihood-ratio test was performed to examine its significance level.

The results indicated that the DRAW × FH effect was statistically significant (p = 0.005). Fig. 11 shows that, in No-DRAW runs, TLX ratings increased when the FH was Near rather than Far; this trend was not observed in DRAW runs.

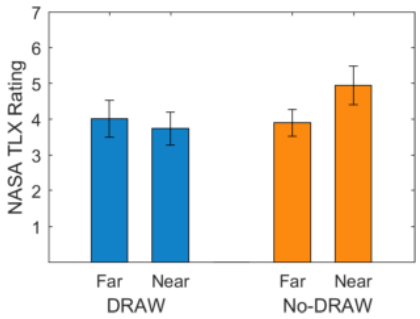

Figure 11. Means and standard errors of controllers' NASA TLX ratings.

E. TMC Workload

The ZTL TMCs reported their real-time workload ratings at 10-minute intervals on a questionnaire (during the blackout periods). The Simplified Subjective Workload Assessment Technique (S-SWAT) scale was used [22]. The final S-SWAT rating was obtained as a simple average of the three subscale ratings without weighting (Time Load, Mental Effort Load, and Psychological Stress Load), each recorded on a continuous scale from 0 to 100, where 0 corresponded with the lowest workload and 100 the highest.

The two ZTL TMCs' S-SWAT rating ranges turned out to differ widely (0 to 5 versus 9 to 60). Considering both TMCs were equally experienced, the large difference between their ratings seemed more likely due to the individuals' perceptual differences, as well as the relatively large SWAT subscale range (0 to 100), rather than the absolute workload level difference. To focus on how their ratings were changed by various factors, their ratings were normalized by mapping their minimum and maximum ratings to 0 and 100, respectively. Note that the normalization was performed on the final S-SWAT ratings after the summation, not on the individual subscale ratings, to preserve the ratio among the subscale ratings.

An LMM analysis with the same model in Table II was applied to the TMCs' normalized S-SWAT ratings. Segments 1-4 were defined as 10 to 30, 40 to 60, 70 to 90, and 100 to 110 scenario-elapsed minutes, respectively. (Note Segment 4 had a shorter period than the other segments, and was typically a low-workload time after the TMC finished implementing all reroutes.) Participant effect represented the two ZTL TMCs.

The likelihood-ratio test detected a significant DRAW × Participant (TMC) effect (p < 0.001)—one ZTL TMC reported higher workload in No-DRAW runs than in DRAW runs, whereas the other reported the opposite. The TMC who reported higher workload in DRAW runs sometimes had problems with the DRAW user interface (e.g., to evaluate a DRAW reroute advisory, the TMC mistakenly opened the current Flight Plan route instead of the advisory route), which may have resulted in higher workload ratings. These issues could be resolved by improving user interface design and/or user training.

The following results were common between the two ZTL TMCs. For one, the DRAW × Segment 3 interaction effect was significant (p = 0.008). Fig. 12 indicates that, in Segment 3 (70 to 90 scenario-elapsed minutes), the TMCs' S-SWAT workload ratings were significantly lower in DRAW runs than in No-

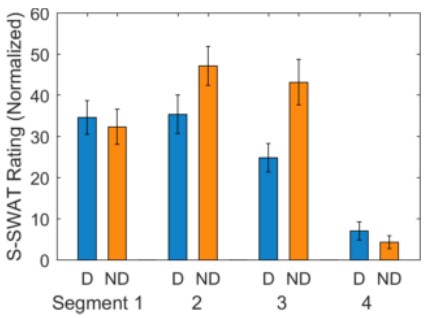

Figure 12. Means and standard errors of TMC S-SWAT ratings by DRAW × Segment. "D" is DRAW runs, and "ND" No-DRAW runs.

DRAW runs. Segment 2 seems to show a similar trend, though it was found only marginally significant (p = 0.070). Their S-SWAT ratings were significantly lower in Segment 4 regardless of DRAW condition (p < 0.001), as also seen in Fig. 12 and consistent with observation.

F. Additional TMC Responses and Feedback

In this study, the 5-minute-on-5-minute-off blackout process allowed the TMCs to work on the traffic only half the time. One of the post-run questions asked the ZTL TMCs what they felt about the time available to them. The responses were collected on a 5-point scale, with 1 being "too short" and 5 "too long." All responses were either 2 ("somewhat too short") or 3 ("neutral"), implying the blackout placed a slight time pressure on them, but not too much.

The ZTL TMCs, who were the direct users of DRAW, commented that the DRAW Advisory List helped lessen their workload. They also said the anticipated weather locations provided by the CWAM WAFs in DRAW runs were helpful. They mentioned that the group reroute advisories in DRAW runs were valuable, especially when workload was high. (The group TP function was actually available in both DRAW and No-DRAW runs, but when the TP was not initiated by a DRAW group advisory, the TMC had to first manually specify the group.)

In this study, the ZJX TMC did not use DRAW directly and only provided the ZTL TMC reroute-planning consultation when needed. The ZTL TMC was designated as the user because the arrival metering was for KATL in ZTL. However, this task-allocation setup did not turn out to be realistic. Both ZJX and ZTL TMCs said that ZJX should have been the DRAW user, as most of the weather-avoidance reroutes occurred in the ZJX airspace. In field operation, the Center in which the weather is located normally initiates the weather-avoidance reroutes. This point raises an interesting question for Center responsibilities in adjacent-Center arrival-metering operations under weather, which will be revisited in the next section.

IV. Discussion

What was clearly demonstrated in this study was that the distance of the FH had significant influence over arrival metering performance—when DRAW was not provided. A FH located far from the MFx resulted in more Flight Plan route amendments after the FH and larger magnitudes of ATA errors with respect to the STA assigned at the FH ($|E_{fh}|$), whereas a FH near the MFx significantly increased the en route sector controller's workload, when DRAW was not provided. However, DRAW use avoided these negative effects and made the ATA accuracy and sector controller workload more robust to the choice of FH distance.

The robustness to the FH distance afforded by using DRAW potentially offers an advantage for advanced metering operations, where multiple FHs have to be set up in a complex manner—such as the FAA's Extended Metering and Coupled Scheduling [29], or the SESAR Extended Arrival Management (E-AMAN) in Europe [30]—since DRAW gives some leeway to system designers in choosing the FH distances. For instance, with DRAW, the number of FHs can be reduced by placing them far apart without compromising the arrival metering delivery performance and the controller workload.

Contrary to the expectation, the study did not find any evidence that the DRAW use resulted in more Flight Plan route amendments before the FH. However, there was some evidence that the DRAW's forecast weather information likely helped reducing the number of route amendments after the FH, especially when the FH was located far from the MFx.

The present study focused more on the ATA error with respect to the STA at FH (E_{fh}) than that with respect to the STA at MFx (E_{mfx}). Other arrival-metering studies, including the previous DRAW study [19], typically focused on the latter, as E_{mfx} is the final metering performance that directly impacts the TRACON work. However, the former may be a more critical indicator from the Center's viewpoint, as a larger E_{fh} likely requires more work from the en route sector controllers and the TMCs to sustain arrival metering. Indeed, in the present study, if only E_{mfx} were looked at, both DRAW and No-DRAW runs managed to achieve acceptable arrival metering performance. Looking at E_{fh}, however, revealed a different story. In addition, the STA at the FH was optimized for interval management and runway sequence, but the STA at MFx, if manually adjusted after the FH, may no longer be optimal depending on the method of the manual adjustment: if the STAs had been adjusted by a TMC's *rippling the list*, the adjusted STAs were derived by the scheduler, and thus, still optimal. If the STAs had been adjusted by a controller's STA *swap* between two airplanes of the same class, then the STAs were probably still optimal. Other cases may result in non-optimal STAs at the MFx. Smaller E_{fh} can reduce the chance of resulting in non-optimal STAs at the MFx.

The study's adjacent-Center metering operational setup also sheds light on inter-Center coordination issues. Operational concepts for DRAW, or any arrival metering involving extended distance under severe weather conditions, need to clarify how to conduct proper coordination when the responsibilities for weather avoidance and arrival metering fall into two different Centers. Simple agreements, such as which arrival transition the adjacent Center should deliver, may be sufficient to some extent. However, to allow more flexible rerouting involving multiple Centers to maximize efficiency, a technological solution to enable the relevant Centers and the Command Center to view, discuss, and modify the reroute may be required.

V. Conclusion

A human-in-the-loop simulation evaluation study of ZTL arrival metering operation under severe weather conditions in coordination with an adjacent ZJX Center demonstrated that DRAW use reduced both the ZTL TMC workload during busy periods, and the number of manual adjustments of frozen STAs implemented by the controllers and TMCs, provided that the weather forecast is reasonably accurate. DRAW use also made the ATA accuracy and the controller workload more robust to the adverse effects of FH distance. This benefit may help in advanced extended-metering operations with the presence of convective en route weather. The results also suggested need for clear strategy for facilitating inter-Center and the Command Center communication and coordination.

Acknowledgment

The authors thank the DRAW team members at NASA Ames—Greg Wong, Chuhan Lee, Kim Hoy, Andrew Biederman, and Matt Blanken—for their hard work that made this complex simulation study possible. Kevin Witzberger and Kapil Sheth of NASA Ames and Matt Modderno of the FAA provided the team with helpful guidance for the project work, and Phil Bassett of the FAA gave valuable insights into the ZTL and ZJX procedures. Last but not least, the authors express great gratitude for the TMC, controller, and pilot participants.

References

[1] H. N. Swenson, T. Hoang, S. Engelland, D. Vincent, T. Sanders, B. Sanford, and K. Heere, "Design and operational evaluation of the Traffic Management Advisor at the Fort Worth Air Route Traffic Control Center," 1st USA/Europe Air Traffic Management R&D Seminar, Saclay, France, June 1997.

[2] N. Hasevoets and P. Conroy, "AMAN status review 2010," ed. 0.1, Brussels: EUROCONTROL, 2010.

[3] N. Hasevoets and P. Conroy, "Arrival Manager: Implementation guidelines and lessons learned," ed. 0.1, Brussels: EUROCONTROL, 2010.

[4] Single European Sky ATM Research Joint Undertaking, "European ATM Master Plan," MG-04-15-465-EN-N, Belgium, ed. 2015.

[5] S. Shresta, B. Hogan, P. Railsback, M. Yankey, J. DeArmon, K. Levin, K. Hatton, and M. Merkle, "Comparison of strategies for continuing Time-Based Metering during inclement weather," *Air Traffic Control Quarterly*, vol. 22, no. 1, pp. 21-47, 2014.

[6] C. Gong and D. McNally, "Dynamic arrival routes: A trajectory-based weather avoidance system for merging arrivals and metering," 15th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Dallas, TX, USA, June 2015.

[7] M. Matthews and R. DeLaura, "Assessment and interpretation of en route weather avoidance fields from the convective weather avoidance model," 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Ft. Worth, TX, USA, September 2010.

[8] S. Campbell, M. Matthews, and R. DeLaura, "Evaluation of the convective weather avoidance model for arrival traffic," 12th AIAA Aviation Technology, Integration, and Operations (ATIO) and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Indianapolis, IN, USA, September 2012.

[9] J. S. DeArmon, H. Gao, D. Chaloux, and M. Hokit, "Metering during severe en route weather via Advanced Flight-Specific Trajectories (AFST)," AIAA Aviation Forum, Denver, CO, USA, June 2017.

[10] D. McNally, K. Sheth, C. Gong, P. Borchers, J. Osborne, D. Keany, B. Scott, S. Smith, S. Sahlman, C. Lee, and J. H. Cheng, "Operational evaluation of Dynamic Weather Routes at American Airlines," 10th USA/Europe Air Traffic Management R&D Seminar, Chicago, IL, USA, June 2013.

[11] D. McNally, K. Sheth, C. Gong, M. Sterenchuk, S. Sahlman, S Hinton, C. H. Lee, and F. T. Shih, "Dynamic weather routes: Two years of operational testing at American Airlines," *Air Traffic Control Quarterly*, vol. 23, no. 1, pp. 55-81, 2015.

[12] K. D. Bilimoria, M. Hayashi, and K. Sheth, "Human-in-the-loop evaluation of Dynamic Multi-Flight Common Route Advisories," AIAA Aviation Forum, Atlanta, GA, USA, June 2018.

[13] J. E. Robinson III, J. Thipphavong, and W. C. Johnson, "Enabling Performance-Based Navigation arrivals," 11th USA/Europe Air Traffic Management R&D Seminar, Lisbon, Portugal, June 2015.

[14] T. Prevot, J. Mercer, J. Homola, S. M. Hunt, A. N. Gomez, N. Bienert, F. G. Omar, J. Kraut, C. L. Brasil, and M. G. Wu, "Arrival metering precision study," AIAA Modeling and Simulation Technologies Conference, Kissimmee, FL, USA, January 2015.

[15] S. Sharma and M. Wynnyk, "Assessment of Delivery Accuracy in an Operational Like Environment," 16th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Washington, DC, USA, June, 2016.

[16] H. F. Vandevenne and J. W. Andrews, "Effects of metering precision and terminal controllability on runway throughput," *Air Traffic Control Quarterly*, vol. 1, no. 3, pp. 277-297, 1993.

[17] J. Thipphavong and D. Mulfinger, "Design considerations for a new terminal area arrival scheduler," 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Ft. Worth, TX, USA, September 2010.

[18] S. Shresta and R. H. Mayer, "Benefits and constraints of time-based metering along RNAV STAR routes," 28th Digital Avionics Systems Conference, Orlando, FL, USA, October 2009.

[19] D. R. Isaacson, M. Hayashi, C. Gong, G. Wong, and H. Tang, "Laboratory evaluation of dynamic routing of air traffic in an en route arrival metering environment," AIAA Aviation Forum, Atlanta, GA, USA, June 2018.

[20] A. Vanwelsenaere, J. Ellerbroek, J. M. Hoekstra, and E. Westerveld, "Effect of popup flights on the Extended Arrival Manager," *Journal of Air Transportation*, vol. 26, no. 2, pp. 60-69, April 2018.

[21] J. E. Robinson III, A. Lee, and C. F. Lai, "Development of a high-fidelity simulation environment for Shadow-Mode Assessments of Air Traffic Concepts," Modeling and Simulation in Air Traffic Management Conference, Royal Aeronautical Society, London, UK, November 2017.

[22] A. Luximon and R. S. Goonetilleke, "Simplified subjective workload assessment technique," *Ergonomics*, vol. 44, no. 3, pp. 229-243, 2001.

[23] S. G. Hart and L. E. Staveland, "Development of a multi-dimensional workload scale: Results of empirical and theoretical research," in P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183), Amsterdam, Netherlands: North-Holland, 1988.

[24] K. C. Hendy, K. M. Hamilton, and L. N. Landry, "Measuring subjective workload: When is one scale better than many?" *Human Factors*, vol. 35, no. 4, pp. 579-601, 1993.

[25] B. T. West, K. B. Welch, and A. T. Galecki, *Linear mixed models: A practical guide using statistical software*, 2nd edition, Boca Raton, FL: CRC Press, 2014.

[26] R Core Team, "R: A language and environment for statistical computing," Vienna, Austria: *R* Foundation for Statistical Computing, http://www.R-project.org/, 2018.

[27] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1-48, 2015.

[28] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1-26, 2017.

[29] B. Lascara, L. A. Weitz, T. Monson, and R. Mount, "Measuring performance of initial Ground-based Interval Management–Spacing (GIM-S) operations," 12th USA/Europe Air Traffic Management R&D Seminar, Seattle, WA, USA, June 2017.

[30] S. Muresean, "AMAN information extension to en-route sectors - concept of operations," ed. 1.0, Brussels: EUROCONTROL, 2009.