

# Human-in-the-Loop Evaluation of Dynamic Multi-Flight Common Route Advisories

Karl D. Bilimoria,<sup>\*</sup> Miwa Hayashi,<sup>†</sup> and Kapil Sheth<sup>‡</sup>  
*NASA Ames Research Center, Moffett Field, CA 94035*

Flights often experience large delays when they are routed around weather. Multi-flight common route advisories provide delay recovery by suggesting time-saving re-routes for groups of flights whose current weather-avoidance routes have become outdated because the weather has dissipated and/or moved away. A laboratory evaluation of these advisories was conducted by four subject matter experts having extensive experience in traffic flow management operations. These experts provided a total of 120 data points in the airspace of Houston Center. The multi-flight common route tool provides time-saving route change advisories taking into account flight plans, wind fields, and the spatio-temporal evolution of predicted convective weather. It is not designed to account for complex operational factors such as non-standard sector traversal and interactions with local traffic management initiatives; hence a relatively low percentage (37%) of advisories generated by the tool were rated as acceptable. However, a high percentage (81%) of advisories were rated as acceptable after the subject matter experts used the tool's user interface to make route modifications that accounted for relevant operational factors not considered by the tool. The workload associated with using the tool, as measured by the NASA Task Load Index, was quite low (1.1 on a scale of 0 to 10). The results of this evaluation make a good case for human-automation teaming to design operationally valid weather re-routes for delay recovery.

## Introduction

This paper presents a follow-on study of the effort reported in [1]; background material from that paper is repeated here for the reader's convenience. Traffic flow management [2] seeks to balance the demand for National Airspace System (NAS) flight resources, such as airspace and airports, with the available capacity. When weather blocks normal air traffic routes, traffic flow managers must re-route affected flights for weather avoidance. Depending on the nature and scope of the weather, traffic flow managers may use pre-coordinated re-routes such as Playbook routes or Coded Departure Routes [2], or may design *ad hoc* local re-routes. The routes of the affected flights are modified accordingly, and the modified routes of these flights will be less efficient than their nominal routes due to increased flight time and fuel burn. In current traffic flow management operations, the transition into a weather avoidance re-routing initiative is typically implemented more aggressively than the transition out of that initiative as the weather dissipates or moves away. For example, strategic large-scale Playbook re-routes are sometimes left in place for hours as initially implemented, even though the weather has changed significantly. There is an opportunity to periodically modify the re-routing plan as the weather evolves, thereby attenuating its adverse impact on flight time and fuel consumption; this is called delay recovery.

Multi-Flight Common Routes (MFCR) is a NASA-developed operational concept for delay recovery, designed to assist traffic flow managers in efficiently updating weather avoidance routes after the original routes have become outdated due to subsequent evolution of the convective weather. In order to reduce the number of advisories that

---

<sup>\*</sup>Aerospace Research Engineer, Flight Trajectory Dynamics & Controls Branch, Mail Stop 210-10; [karl.bilimoria@nasa.gov](mailto:karl.bilimoria@nasa.gov). Fellow, AIAA.

<sup>†</sup>Aerospace Engineer, Aerospace High Density Operations Branch, Mail Stop 210-8; [miwa.hayashi@nasa.gov](mailto:miwa.hayashi@nasa.gov). Senior Member, AIAA.

<sup>‡</sup>Manager, Airspace Technology Demonstration-3, Flight Trajectory Dynamics & Controls Branch, Mail Stop 210-10; [kapil.sheth@nasa.gov](mailto:kapil.sheth@nasa.gov). Associate Fellow, AIAA.

need to be evaluated and coordinated, the MFCR tool groups multiple flights and merges them on a common route segment to provide a new flow of re-routed traffic.

The MFCR concept of operations can be summarized as follows. The MFCR tool generates time-saving re-route advisories for a group of flights that are currently airborne within the same Air Route Traffic Control Center. This advisory is presented to a Traffic Management Coordinator (TMC) in that Center, who evaluates it and has the option to modify it using the MFCR tool's graphical user interface (GUI). If the TMC finds the advisory to be operationally appropriate, he/she would coordinate with the Area Supervisor(s) of the sectors that currently control the flights in the advisory as well as TMCs in other facilities that would be affected by the advisory; coordination with the Air Traffic Control System Command Center (ATCSCC) may also be necessary for multi-Center routings. When the TMC accepts the finalized MFCR advisory via the GUI, the corresponding flight plan amendments are sent to the air traffic displays of the appropriate sector controllers, using the AirBorne ReRoute (ABRR) capability that is already operational in the NAS. The sector controllers then offer time-saving route modifications to the pilots of the affected flights via datalink or voice communications, and then implement the corresponding flight plan amendment for the pilots who accept it (after consultation with their dispatcher if necessary).

MFCR is implemented as an application in the software environment of the Future Air traffic management Concepts Evaluation Tool (FACET) [3]. This application is the NAS Constraint Evaluation and Notification Tool (NASCENT). An assessment of NASCENT's single-flight re-routing functionality was presented in [4]. An analysis of a preliminary version of NASCENT's multi-flight functionality was presented in [5]. The present paper reports a human-in-the-loop (HITL) evaluation of MFCR, covering the operational concept, functionalities, algorithm, and GUI, with emphasis on the dynamic nature of MFCR advisories.

### Features of MFCR Tool

The MFCR tool suggests wind-corrected time-saving weather-avoidance routes for groups of flights that are currently airborne within the same Center and have on their current flight plan a common downstream waypoint, called the Return Capture Fix. The routes of all flights in a group are merged at a waypoint called the Merge Point, and the flights then fly along a common route segment to the Return Capture Fix after which they resume their original routing.

Figure 1 presents an example MFCR advisory in the airspace of Houston Center (ZHU). The green routes depict the active flight plan which corresponds to an implementation of Playbook "LEV West Partial." This routing is typically used when a large weather system in the south-central U.S. blocks transcontinental traffic. This Playbook re-route merges west-bound traffic from the southeastern U.S. at fix THX (Three Rivers, TX) and sends the traffic via some intermediate waypoints to the fix ELP (El Paso, TX). After ELP, traffic follows standard jet routes to their destinations in the western U.S. In Fig. 1 this Playbook routing is still in effect even though the weather (depicted by yellow, orange, and red contours in the large map on the left) has moved to the north. This presents an opportunity for MFCR to advise an updated set of time-saving weather-avoidance routes, shown by yellow lines.

The MFCR advisory shown in Fig. 1 identifies nine flights in ZHU that are advised to be merged at fix LEJON (the Merge Point) and then sent direct to ELP (the Return Capture Fix) with the rest of the route unchanged. The new routes of these nine flights are estimated to provide a total time savings of 53 min (an average of about 6 min per flight). Also, these flights have been organized into a new flow along the common route segment from LEJON to ELP. This MFCR advisory provides a beneficial intermediate routing solution prior to entirely lifting the Playbook restriction when the weather dissipates fully. If the new flow continues to be free of weather, there is also an opportunity to update the existing Playbook routing and provide time savings to other aircraft that are currently in Centers upstream of this route modification.

The right-hand side of Fig. 1 shows forecasted sector loading for the current (Playbook) routing as well as the MFCR-advised routing. Red or yellow sectors indicate a prediction of traffic demand in excess of sector capacity, with red/yellow indicating a higher/lower level of confidence in the prediction of sector overload; the degree of overload is also indicated by text. Although the MFCR algorithm does not currently utilize sector congestion information to compute re-route advisories, it is presented to the traffic manager for informational purposes. The traffic manager may use this information to modify the MFCR advisory for operational acceptability by avoiding red/yellow sectors. For the scenario shown in Fig. 1, the MFCR routing does not result in any additional red or yellow sectors relative to the Playbook routing.

MFCR advisories are updated every minute and therefore may change over time: some flights may be dropped from the group, and/or new flights may be added to the group, and/or the structure of the advised route may change. In our previous Subject Matter Expert (SME) evaluation of MFCR advisories, reported in [1], the dynamic aspect

was suppressed and the SMEs were presented with a traffic scenario that was effectively frozen in time. All traffic, including flights in the MFCR advisory, remained static. However, the SMEs were given the ability to modify the MFCR advisory and immediately see the effects of their trial plans without their modifications being confounded by the dynamics of the advisory. This enabled the evaluation of an advisory at a single pre-determined instant of time, sufficient for gaining some fundamental insights appropriate for an initial evaluation of MFCR advisories.

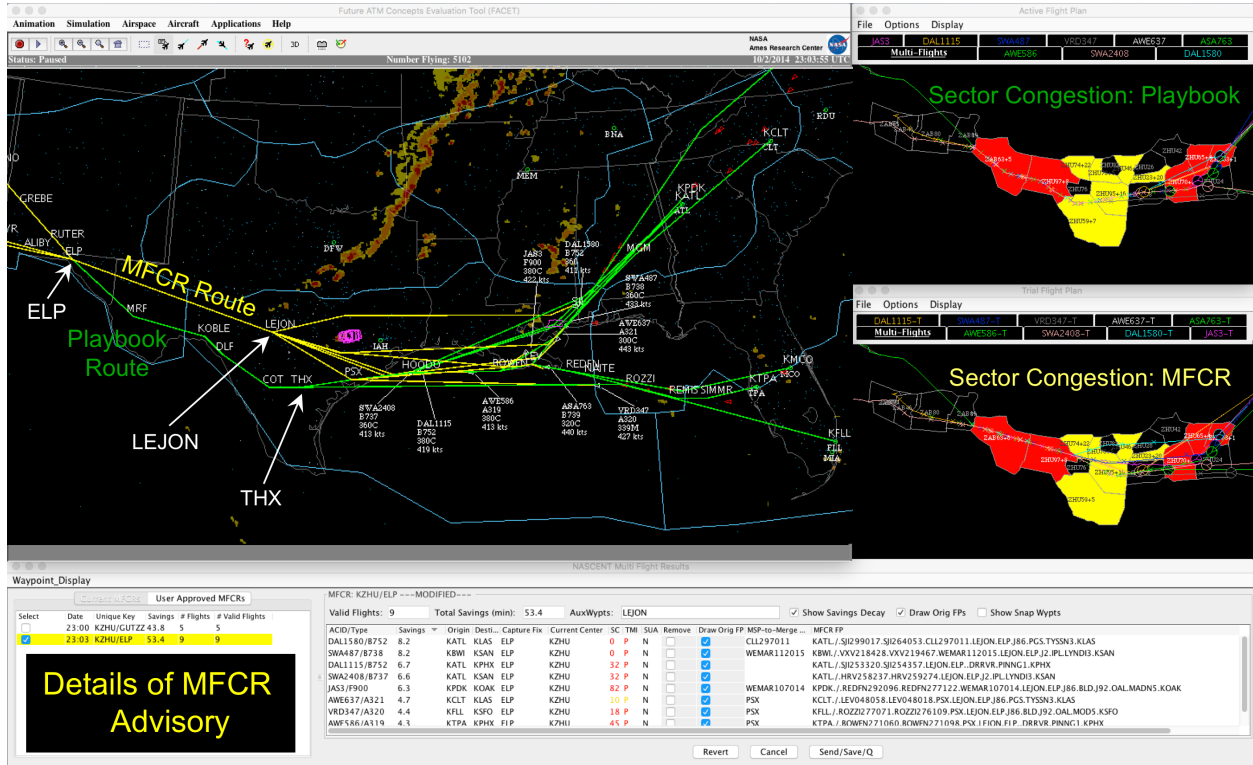


Fig. 1. Example MFCR advisory

The HITL evaluation reported in this paper focused on the dynamic nature of MFCR advisories. Since an advisory could change while the user was in the process of modifying it, the MFCR tool was enhanced with new functionality to accommodate these dynamics. For example, the user was provided with a graphical depiction of the estimated savings decay (not shown in Fig. 1) that would occur in an advisory as time progressed and flights currently in the advisory group moved along their active routes. Additionally, there was an indication (aircraft icons highlighted in red) of upstream Center flights that may soon join the existing MFCR group in ZHU; for example, see Atlanta Center airspace in Fig. 1. The above two functionalities provided the user with the means to assess the tradeoff between immediately implementing the advisory and waiting a few minutes for additional flights.

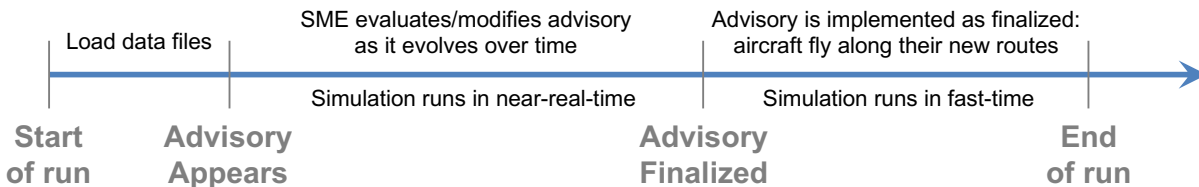
### HITL Evaluation of MFCR

A HITL evaluation of MFCR advisories was conducted in the Air Traffic Control Laboratory at NASA Ames Research Center, 7 – 9 March 2017. The objective of this activity was to evaluate various aspects of dynamic MFCR advisories. Four retired TMCs from Houston Center participated in the experiment as SMEs. Their experience in ZHU Traffic Management Unit operations ranged from 9 to 14 years, with an average of 11.0 years. Two of the participants had supervisory TMC experience. All participants had either retired or departed from the ZHU TMU between 4 to 8 years prior to this HITL evaluation. Two of the four participants had participated in our previous MFCR study [1], and therefore had prior exposure to an earlier version of the MFCR tool. The other two participants did not have any prior exposure to the MFCR tool.

## Data Collection

All participants conducted all parts of the evaluation at the same time. On the morning of the first day, the participants received a comprehensive briefing on MFCR, including its concept of operations, functionalities, algorithm, and GUI. This briefing was followed by a demonstration of the MFCR tool and a tutorial on its various functionalities. On the afternoon of the first day, the participants engaged in a hands-on training activity to exercise all relevant functionalities of the MFCR tool using 10 traffic scenarios selected for the practice session. The entire second day and the morning of the third day were devoted to data collection where each participant was presented with a total of 30 MFCR re-route advisories for evaluation and possible modification using the tool's trial planning functionality; the participants conducted their evaluations independently. This resulted in a total of 120 data points. Each data collection run focused on the evaluation of a single re-routing scenario featuring a MFCR advisory. An observer shadowed each participant to take research notes and answer any questions about the tool's functionalities. After each data collection run was completed, the participants filled out a post-run questionnaire. In the afternoon of the third day, after all data collection runs had been completed, the participants filled out a post-evaluation questionnaire, and then participated in a two-hour debrief session with the researchers, discussing various aspects of MFCR's operational concept, functionalities, algorithm, and GUI.

Figure 2 depicts the timeline (~15 min) of a nominal data collection run for a typical scenario. The process began with the loading of all data files associated with the scenario. After the data files were loaded, the simulation built up the various panels of the MFCR tool's display. When the display buildup was complete, the advisory appeared on the GUI, and the participants began their evaluation of a dynamic re-route advisory as the simulation ran in near-real-time. Participants modified the MFCR-issued advisory route as appropriate, receiving immediate feedback on the resulting changes in flight time savings, sector loadings, and any proximity/penetration of weather polygons. There was a single-click option for reversion to the current MFCR-generated advisory. After the participants completed their evaluation and possible modification (no time limit was imposed), they clicked a button indicating that they had finalized the advisory. The advisory was then implemented as finalized, and simulated aircraft flew along their new routes in fast-time so that the participants could quickly visualize the consequences of implementing their finalized advisory.



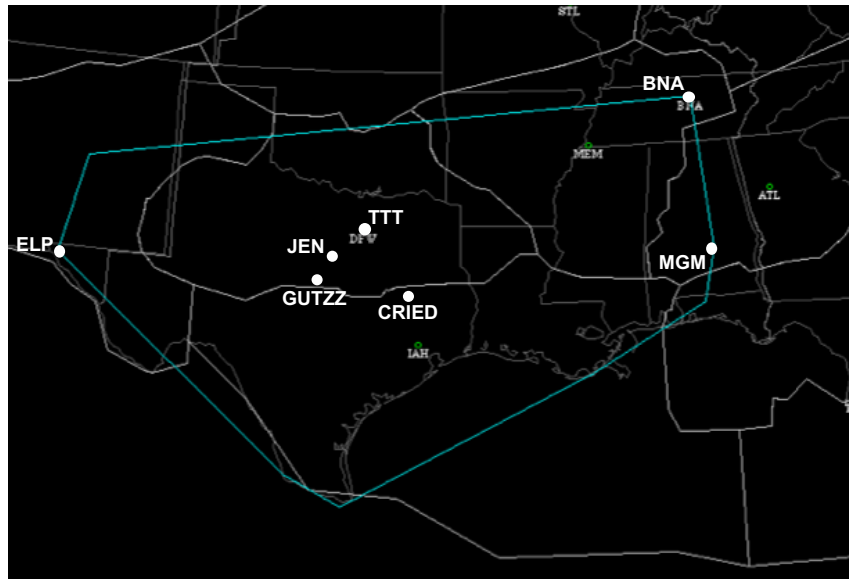
**Fig. 2. Timeline along a nominal 15-minute data collection run**

## Scenarios

The scenarios used in this study featured MFCR advisories for flights in ZHU airspace. Scenarios were generated using actual traffic and weather data for 60 bad-weather days in 2014 and 2015. Traffic data were obtained from the FAA's Aircraft Situation Display to Industry (ASDI) feed, wind data were obtained from the National Oceanic and Atmospheric Administration, and Corridor Integrated Weather System (CIWS) data were obtained from the Massachusetts Institute of Technology – Lincoln Laboratory. The CIWS data were input to the Convective Weather Avoidance Model (CWAM) [6] which utilized pilot deviation probabilities to generate polygons for weather avoidance as a function of altitude and time. Using these data, MFCR software was run in playback mode to generate re-route advisories for flights in ZHU airspace. An advisory was required to have the following properties: a minimum of two flights per advisory, a minimum of three minutes time savings per flight, and a minimum of 10 minutes time savings summed over all flights in the advisory. Details of MFCR advisories were output for an offline analysis that resulted in the selection of the day and time-interval for each scenario; the selection process attempted to provide a diverse set of advisories. The corresponding traffic and weather data files were used to run the scenarios for the evaluation.

Figure 3 shows the Return Capture Fixes for MFCR advisories in the 30 data collection scenarios, along with state (thin grey) and Center (thick grey) boundaries. The blue polygon is Houston Center's limit polygon, constructed for research use from historical traffic data; it indicates the furthest downstream route clearances typically issued by ZHU controllers in actual operations. For ZHU flights landing beyond this limit polygon, the

Return Capture Fix is the flight plan's last fix in or on the limit polygon. For ZHU flights landing within this limit polygon, the Return Capture fix is the transition fix for the Standard Arrival Route (STAR) on the flight plan (if the STAR is missing in the flight plan, the last fix before the destination airport is used instead). ELP was the Return Capture Fix for 14 scenarios; TTT for five; GUTZZ for four; BNA, CRIED, and MGM for two each; JEN for one. For the CRIED, GUTZZ, JEN, and TTT scenarios, the MFCR advisories featured flights going to Dallas – Fort Worth airport (DFW). The ELP scenarios featured West-bound flights in their advisories, while the BNA and MGM scenarios featured East-bound flights in their advisories.



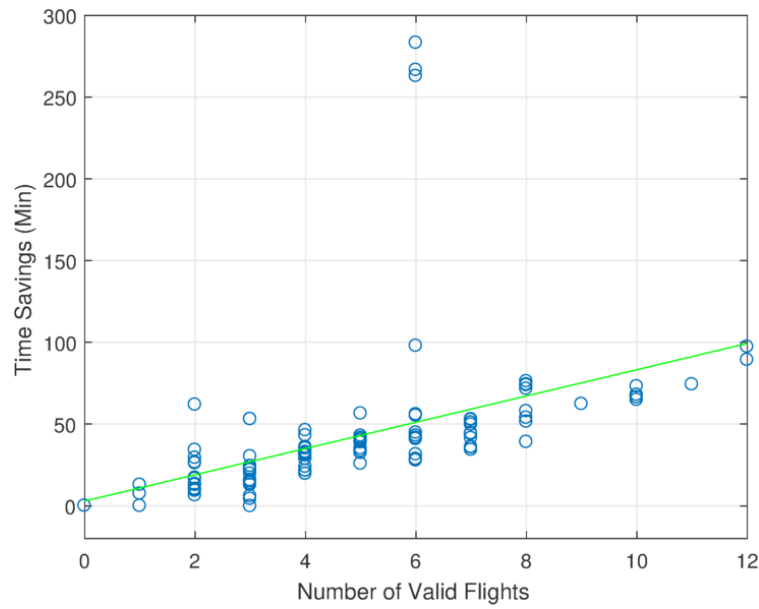
**Fig. 3. Return Capture Fixes**

### Evaluation Results

The participants evaluated advisories as originally presented by the MFCR tool, and modified those that needed adjustment for operational factors not addressed by the tool. In this work, the re-route advisory generated by the MFCR tool is called the Initial Advisory with the understanding that this advisory is dynamic (one-minute update rate) and hence may change over time. The end result of any sequence of participant-generated modification(s) of the Initial Advisory is called the Post-Modification Advisory. For the purposes of analysis, the Final Advisory is defined as the Initial Advisory if no net modification was made, or the Post-Modification Advisory if any net modification was made.

The scenarios featured various numbers of flights in their MFCR advisories. Due to the dynamic nature of the scenarios and the variability of MFCR tool usage across participants, the number of flights in a Final Advisory showed some variation across participants (for the same scenario). Figure 4 shows the total flight time savings for each Final Advisory as a function of number of valid (no weather penetration) flights in the finalized advisory. As expected, the total savings increased with number of flights; the least-squares regression line is shown in green. The extreme outliers correspond to a scenario where a large-savings shortcut was available.

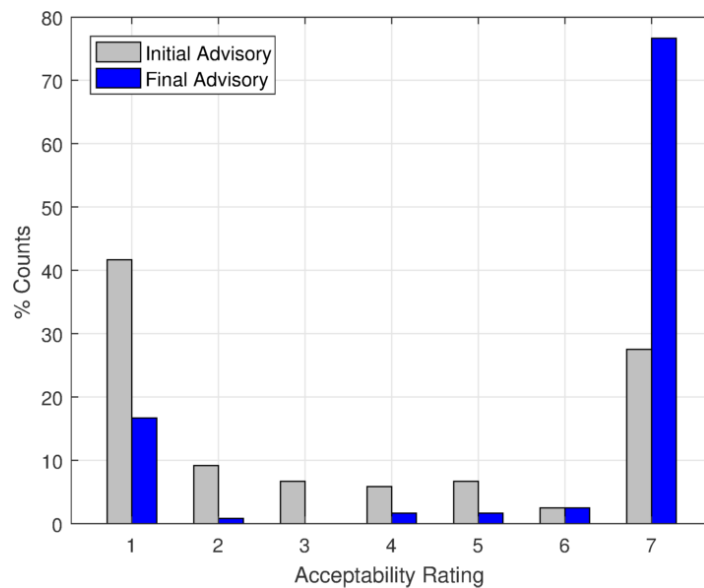
The post-run and post-evaluation questionnaires collected various types of data, including participants' commentary and ratings on several aspects of the MFCR advisory. To examine what effects influenced the various ratings, a Linear Mixed Model repeated-measures regression analysis was conducted [7]. The statistical software *R* and its packages *lme4* and *lmerTest* were used for the analysis [8 – 10]. A regression model was built to examine the participants' ratings of the MFCR advisories for statistically significant variations across the following effects: Participant, Run Index Number, Return Capture Fix, Time Savings, and Number of Auxiliary Waypoints (between Merge Point and Return Capture Fix) in the Final Advisory. Participant was treated as a random effect, and all others were treated as fixed effects. Return Capture Fix and Participant effects were categorical variables, and all others were continuous variables. Participant effects are not covered here because variations in human-subject ratings are to be expected. The other effects are covered, when relevant, in the discussion of the various ratings.



**Fig. 4. Flight time savings vs. number of valid flights**

#### *Acceptability of MFCR Advisories*

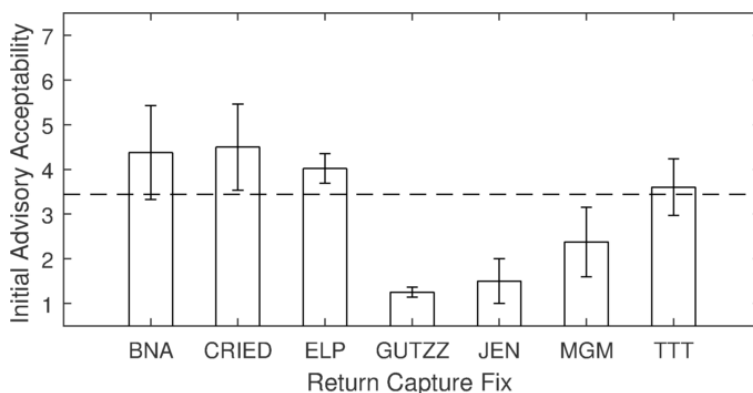
The post-run questionnaire collected ratings of the acceptability of each Initial Advisory and Final Advisory. To assess the Initial Advisory, the participants rated their level of agreement with the statement: “The MFCR re-route advisory as originally presented to you was acceptable.” Responses were collected on a 7-point Likert scale, with 1 = Disagree (lowest acceptability), 4 = Neutral, and 7 = Agree (highest acceptability). Similarly, the participants also rated the acceptability of the Final Advisory in response to the statement: “The MFCR advisory as finally modified by you was acceptable.” If the participant did not modify the Initial Advisory, the latter statement was not presented to the participant, and the acceptability rating for the Final Advisory was defined to be the same as that for the Initial Advisory.



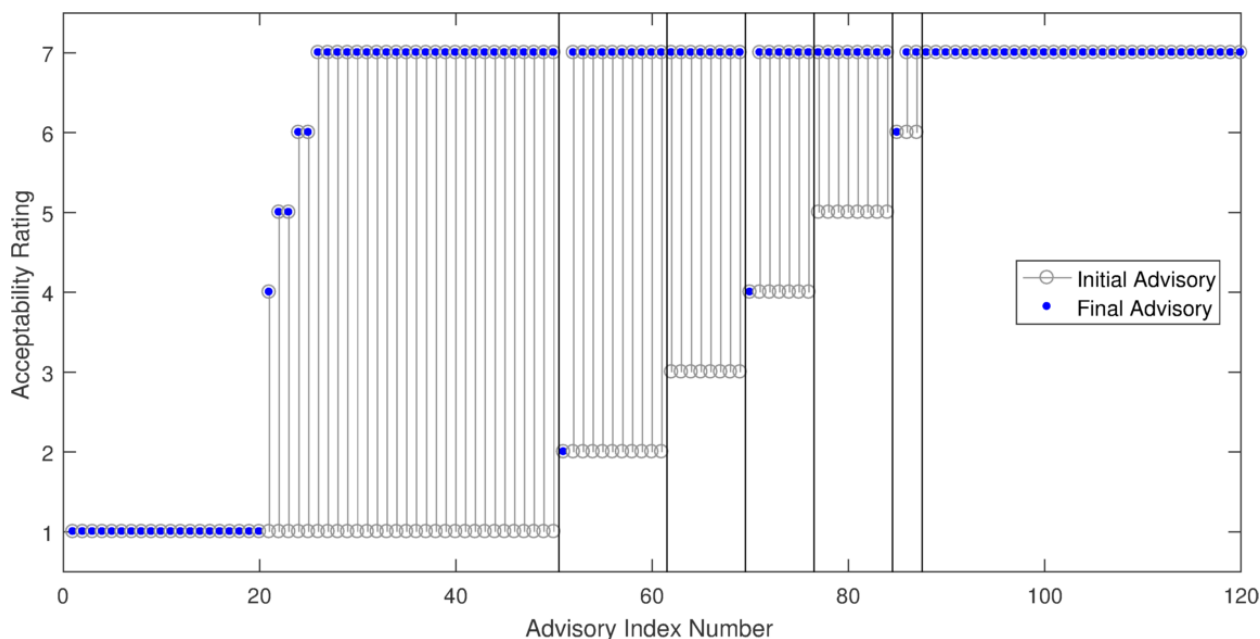
**Fig. 5. Distribution of acceptability ratings for Initial and Final Advisories**

Figure 5 presents the distribution of acceptability ratings for Initial and Final Advisories. In this analysis, MFCR advisories with ratings of 5, 6, or 7 are considered to be “acceptable.” Out of a total of 120 data points, 37% had acceptable ratings for advisories as originally presented by the MFCR tool, and 81% had acceptable ratings for advisories as finally modified by the participants.

Scenarios with the same Return Capture Fix tended to present advisories of a similar route structure. Figure 6 presents average Initial Advisory ratings broken out by Return Capture Fix; the dashed line indicates the overall average. The regression analysis showed, with greater than 95% confidence, that the Initial Advisory acceptability ratings were lower when the Return Capture Fix was either GUTZZ or JEN. Out of 30 scenarios, there were four scenarios with GUTZZ and one scenario with JEN as the Return Capture Fix. Both GUTZZ and JEN are located to the southwest of DFW airport (see Fig. 3). In all of the GUTZZ or JEN scenarios, the flights in the advisory were approaching DFW from the southeast while a line of weather blocked the southeast approach to DFW. For these scenarios, the participants explained that they gave low acceptability ratings because the Initial Advisories would: (1) put arrival flights on routes crossing the departure flow (i.e., the Southbound departures from DFW), (2) reduce the distance available for Fort Worth Center (ZFW) sector controllers to sequence the arrivals, and, (3) cause a large turn at the ZFW/ZHU Center boundary.



**Fig. 6. Initial acceptability ratings vs. return capture fix**



**Fig. 7. Relationships between acceptability ratings of Initial Advisories and Final Advisories**



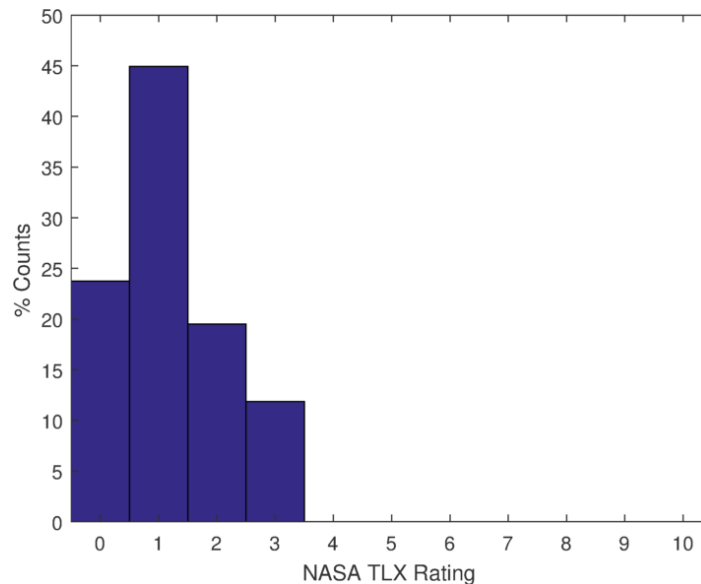
All of the 120 individual ratings for Initial and Final Advisories are presented in Fig. 7, with data points sorted by acceptability rating to provide additional insight. For example, 11 of the Initial Advisories received a rating of 2, indicating a low level of acceptability. In one of those cases, the Initial Advisory received no modification; hence the Final Advisory was assigned the same rating of 2. In the other 10 cases, the participants' modifications greatly improved the ratings of the Final Advisories to 7 (highest acceptability). The rest of the data (with the notable exception of data with Initial Advisory ratings of 1) show a similar trend, with almost all Final Advisories receiving ratings of 7. This indicates that the MFCR tool provides the user (TMC) with the functionality to satisfactorily modify the Initial Advisories to incorporate complex operational aspects not considered by the algorithm. Some of these operational aspects are presented at the end of this section (see Key Insights).

### **Workload**

NASA Task Load Index (TLX) ratings [11] were collected after each run to assess the workload level during the run. The participants recorded their self-assessed rating for each of the six NASA TLX subscales (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) on a rating scale where 0 represented the lowest Demand/Effort/Frustration, and 10 the highest; for the Performance category, 0 represented the poorest performance, and 10 the best. After the Performance ratings were reversed to be consistent with the other subscales' directions, the final NASA TLX rating was obtained as the simple unweighted average of the six subscale scores [12]. The simulation environment used for the MFCR evaluation did not model the baseline tasks performed by TMCs. Hence the workload ratings obtained were part-task in nature, reflecting only the tasks associated with exercising the various functionalities of the MFCR tool.

Figure 8 presents the distribution of NASA TLX ratings. Two of the 120 data points were excluded from the plot due to missing data in their subscale ratings. Figure 8 shows that all the workload ratings are clustered on the lower end of the NASA TLX rating scale, ranging from 0 to 3.33 with an average value of 1.06, indicating that the MFCR-related workload levels were quite low.

The regression analysis showed, with greater than 99% confidence, that the NASA TLX ratings decreased by 0.02 as the Run Index Number increased by one. This indicates a learning effect, i.e., as the participants performed more data runs, they felt that the overall task became easier to perform. The regression analysis also showed that the NASA TLX ratings were higher when the Return Capture Fix was either MGM or CRIED.

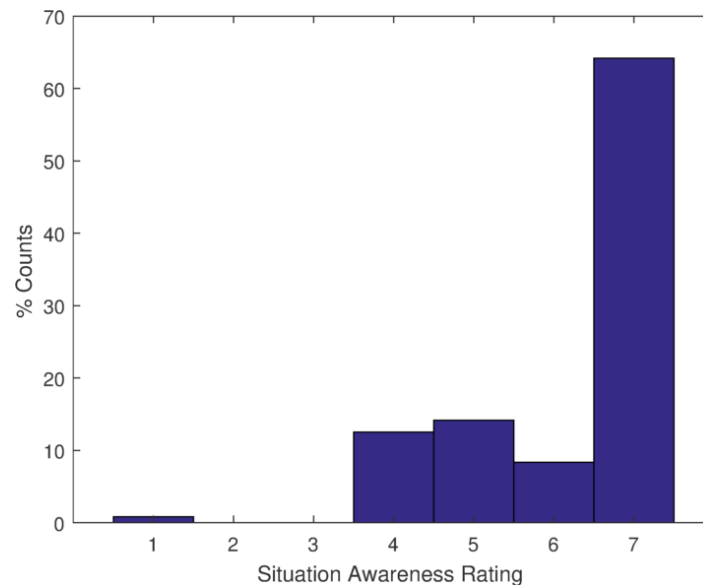


**Fig. 8. Distribution of NASA TLX ratings**



### ***Situation Awareness***

In order to gauge the participants' general perception of their situation awareness level, they were asked after each run to rate their level of agreement with the statement: "I had all the necessary information to make the right decision." Responses were collected on a 7-point Likert scale, with 1 = Disagree (lowest situation awareness), 4 = Neutral, and 7 = Agree (highest situation awareness). Figure 9 plots the distribution of these ratings. The plot shows that virtually all of the responses range from 4 (neutral) to 7 (highest situation awareness). Over 70% of the ratings were either 6 or 7, corresponding to a high level of situation awareness.



**Fig. 9. Distribution of situation awareness ratings**

In the statement used to elicit ratings of situation awareness, the terms "necessary information" and "right decision" were not defined and intentionally left open to the participants' interpretation. A subsequent question (optional) asked what information, necessary to make the right re-route decision, was missing. The responses showed that, in a few situations, the participants wanted information about any miles-in-trail (MIT) restrictions in effect in the area, and any other re-route advisories that may interfere with the MFCR advisories, such as opposite-direction flights on Playbook routes in the same local airspace.

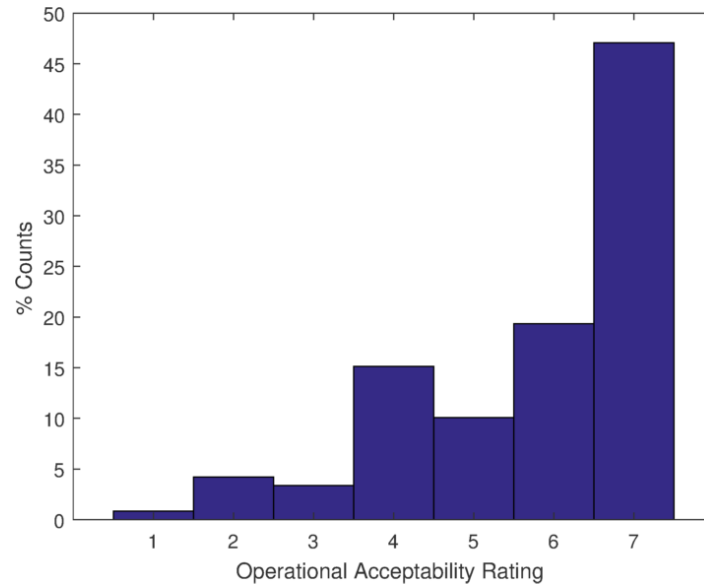
### ***Acceptability of MFCR Operations***

In addition to the acceptability of MFCR advisories (presented earlier), the MFCR operational acceptability rating for each scenario was collected after each run. This pertains to the overall acceptability of traffic flow management operations using the MFCR tool, rather than the acceptability of just the advisory itself. In order to gauge the participants' perception of MFCR operational acceptability, they were asked to rate their level of agreement with the statement: "MFCR operations, as it was experienced in the last run, was acceptable." Responses were collected on a 7-point Likert scale, with 1 = Disagree (lowest acceptability), 4 = Neutral, and 7 = Agree (highest acceptability). Figure 10 shows the distribution of the ratings. One of the 120 data points was excluded from the plot due to missing data. The plot shows that 66% of the ratings were either 6 or 7, corresponding to a high level of operational acceptability.

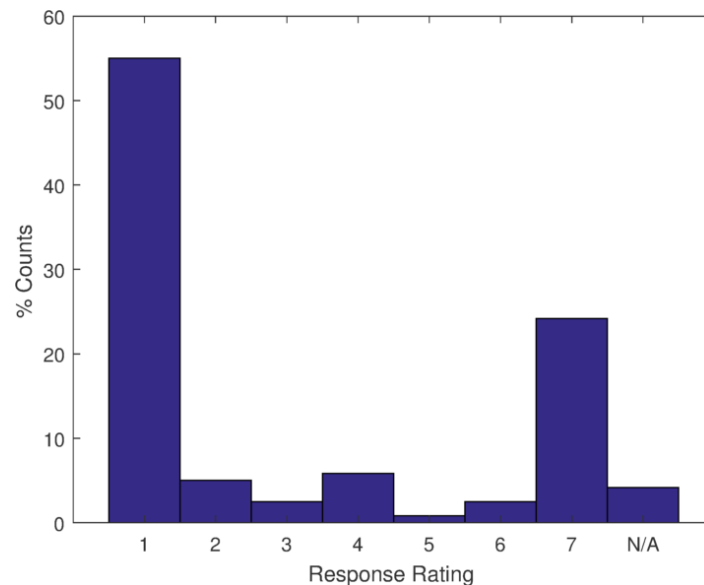
### ***Strategy for Implementation of Dynamic Advisories***

Due to the dynamic nature of MFCR advisories, as times passes, some flights may be dropped from the group, and/or new flights may be added to the group, and/or the structure of the advised route may change from one minute to the next. Two features were provided to help the user (TMC) determine the appropriate time for implementation of these dynamic advisories: (1) a graph showing an estimate of how the group's total time savings would decay as time progressed and flights currently in the advisory group moved along their active routes, and, (2) an indication (aircraft icons highlighted in red, see Fig. 1) of upstream Center flights that may join the existing advisory group

after entering ZHU airspace. For example, if several new flights are expected to join the advisory group in the next 5 minutes, and the savings decay over the next 5 minutes (for flights currently in the advisory group) is small, one option is to wait 5 minutes before implementing the advisory in order to capture a larger amount of total savings in a single advisory. Another option is to immediately implement the current advisory, and then implement another advisory approximately 5 minutes later.



**Fig. 10. Distribution of operational acceptability ratings**



**Fig. 11. Distribution of ratings for advisory implementation strategy**

In order to gain insight into the participants' strategy for implementing dynamic advisories, they were asked after each run to rate their level of agreement with the statement: "In an operational setting, I would have issued the advisory a few minutes later, so that red highlighted flights just outside the ZHU boundary could join the current group of MFCR flights." Responses were collected on a 7-point Likert scale, with 1 = Disagree (would not wait),

4 = Neutral, and 7 = Agree (would wait a few minutes). Figure 11 plots the distribution of the ratings. Not all scenarios had red highlighted flights; hence if the participant did not recall seeing any red highlighted flights, the rating was marked as “N/A.” About 55% of the data points had a rating of 1, indicating a strong preference to implement the advisory right away rather than wait for a few minutes to implement a potentially larger advisory. On the other hand, about 25% of the data points had a rating of 7, indicating a preference to wait for a few minutes to implement a potentially larger advisory.

The regression analysis showed, with greater than 95% confidence, that ratings were higher (stronger preference to wait a few minutes to implement a potentially larger advisory) when the Return Capture Fix was ELP. A possible explanation is that large flows along a Playbook route were feeding the MFCR advisories for ELP scenarios.

### ***System Usability***

A post-evaluation questionnaire was administered after all runs had been completed. This questionnaire included usability ratings elicited by the System Usability Scale (SUS) survey. The SUS, originally designed to quickly measure perceived usability of computer systems [13], has been widely used by researchers over the past 25 years. The SUS score is computed by combining responses to 10 aspects of usability rated on a Likert scale. According to the literature [14], SUS scores from 50 to 70 are considered “Marginal” and scores from 70 to 100 are considered “Acceptable.” The SUS scores of the four participants ranged from 82.5 to 100.0, with a mean score of 93.1, indicating good usability of the MFCR tool.

### ***Key Insights***

Some key insights were obtained from the debrief session and participants’ comments in the post-run and post-evaluation questionnaires. The participants indicated that the MFCR tool identified many time-saving re-routing opportunities that would be difficult to identify manually during air traffic operations in bad weather conditions. Some advisories were found to be operationally acceptable as originally presented by the tool, while others needed modification for operational reasons.

A major reason for modifying Initial Advisories was undesirable sector-traversal features. For example, an advised route sometimes ran for many miles in close proximity to a sector boundary. In some other cases, the route had a very small dwell time in one or more sectors (cutting corners). In yet other cases, the route crossed busy arrival/departure sectors. These types of route attributes would create undesirable complexity and workload for the corresponding sector controllers. The MFCR tool’s “historically used routes” feature, which attempts to compensate for sector-traversal issues by presenting often-used alternatives for the common route segment (when available), was well received by the participants. Another reason for modifying Initial Advisories was to account for interactions with local traffic management initiatives, such as miles-in-trail restrictions for merging streams of traffic in busy airspace. In most cases, the participants were able to quickly resolve any operational issues with minor route modifications using MFCR’s graphical user interface.

While the Initial Advisory routes typically did not penetrate weather regions, the routes sometimes ran on the front side of the motion of a large weather system. Even though the advised routes had a substantial buffer distance from the weather front, the participants were often uncomfortable with this situation due to lack of full confidence in the weather forecast – they preferred routes that ran on the back side of the weather motion. This resulted in low acceptability ratings for Final Advisories in cases where it was not possible to modify the routes to run on the back side of the weather.

## **Conclusions**

MFCR advisories provide time-saving re-routes for groups of flights whose current weather-avoidance routes have become outdated because the weather has dissipated and/or moved away. An evaluation of dynamic MFCR advisories was conducted by four experienced traffic flow managers, in the simulated airspace of Houston Center.

The average acceptability rate of Initial Advisories was 37%, and the average acceptability rate of Final Advisories was 81%. In general, the improvement in acceptability arose from the participants’ modifications that accounted for complex operational factors not considered by the MFCR tool’s algorithm, such as sector traversal properties and interactions with local traffic management initiatives. However, these modifications, made via the MFCR tool’s GUI did not take much effort – the workload for MFCR-related tasks, as measured by the NASA Task Load Index, averaged 1.06 on a scale of 0 to 10. The tool also provided a high level of situation awareness for over 70% of advisories. The overall acceptability of traffic flow management operations using the MFCR tool was rated

as high for 66% of the advisories. The SUS ratings, administered post-evaluation, indicated good usability of the MFCR tool.

Overall, the SMEs were very positive about the MFCR tool's capabilities and its overarching concept of operations. Although difficult for human operators, the automation was able to quickly identify multi-flight route changes for delay recovery, taking into account flight plans, wind fields, and the spatio-temporal evolution of predicted convective weather. Conversely, the human operator was able to quickly adjust this re-route advisory to account for complex operational factors that would be difficult to accurately and comprehensively program into the automation. These results make a good case for human-automation teaming to design valid weather re-routes for delay recovery.

### Acknowledgments

The authors thank the following individuals, all at NASA Ames, for their contributions to the MFCR HITL evaluation: Matt Blanken, Estela Buchmann, Fay Chinn, Alexis Clymer, Paul Cobb, Kaj Edholm, Saugata Guha, Sebastian Gutierrez Nolasco, Nguyen Quach, Scott Sahlman, Mohan Shah, and Fu-Tai Shih.

### References

1. Bilimoria, K.D., Hayashi, M., and Sheth, K., "Subject Matter Expert Evaluation of Multi-Flight Common Route Advisories," Paper No. 2017-3426, *AIAA Aviation Technology, Integration, and Operations Conference*, June 2017.
2. *Traffic Flow Management in the National Airspace System*, Publication 2009-AJN-251, Federal Aviation Administration, Washington, D.C., October 2009.
3. Bilimoria, K.D., Sridhar, B., Chatterji, G., Sheth, K.S., and Grabbe, S., "FACET: Future ATM Concepts Evaluation Tool," *Air Traffic Control Quarterly*, Vol. 9, No. 1, 2001, pp. 1–20.
4. Sheth, K., McNally, D., Somersall, P., Morando, A., Clymer, A., and Shih, F.-T., "Assessment of a National Airspace System Airborne Rerouting Tool," *11<sup>th</sup> USA/Europe Air Traffic Management R&D Seminar*, June 2015.
5. Sheth, K., Clymer, A., Morando, A., and Shih, F.-T. "Analysis of Multi-Flight Common Routes for Traffic Flow Management," Paper No. 2016-4207, *AIAA Aviation Technology, Integration, and Operations Conference*, June 2016.
6. Matthews, M. and DeLaura, R., "Assessment and Interpretation of En Route Weather Avoidance Fields from the Convective Weather Avoidance Model," Paper No. 2010-9160, *AIAA Aviation Technology, Integration, and Operations Conference*, September 2010.
7. West, B., Welch, K.B., and Galecki, A.T., *Linear Mixed Models: Practical Guide Using Statistical Software*, 2nd edition, CRC Press, Boca Raton, Florida, 2014.
8. R Core Team, "R: A language and environment for statistical computing," version 3.4.2, R Foundation for Statistical Computing, Vienna, Austria, 2017. <https://www.R-project.org>.
9. Bates, D., Maechler, M., Bolker, B., and Walker, S., "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, Vol. 6, No. 1, 2015, pp. 1-48.
10. Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B., "lmerTest: Tests in Linear Mixed Effects Models," R package version 2.0-33, 2016. <https://CRAN.R-project.org/package=lmerTest>.
11. Hart, S.G. and Staveland, L.E., "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*, edited by P. A. Hancock and N. Meshkati, North Holland Press, Amsterdam, The Netherlands, 1988, pp. 139-183.
12. Hendy, K.C., Hamilton, K.M., and Landry, L.N., "Measuring Subjective Workload: When is One Scale Better Than Many?" *Human Factors*, Vol. 35, No. 4, 1993, pp. 579-601.
13. Brooke, J., "SUS: A Retrospective," *Journal of Usability Studies*, Vol. 8, No. 2, 2013, pp. 29–40.
14. Bangor, A., Kortum, P., and Miller, J., "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale," *Journal of Usability Studies*, Vol. 4, No. 3, 2009, pp. 114–123.