Proposed Scenarios for the Standardization of the Evaluation of New ATC Technologies and Procedures

Jerry M. Crutchfield
Federal Aviation Administration
Oklahoma City, Oklahoma
Angel Millan
Cherokee CRC, LLC
Oklahoma City, Oklahoma

This project provides Federal Aviation Administration acquisition program managers and system development integration contractors with a standard set of human-in-the-loop simulation scenarios against which new Air Traffic Control (ATC) technologies and procedures can be evaluated. We scripted 24 scenarios, eight scenarios for each of three different types of airspace, including a TRACON arrival sector and both low and high altitude en route sectors. The scenarios were scripted to re-create real world airspaces that analyses showed are associated with complex traffic situations. They included representations of severe weather and high traffic load for the purpose of demonstrating the performance of new ATC technologies and procedures when challenged by such real world events. The scenarios were vetted by retired controllers who had experience working the selected sectors and were provided in a format that allows for input into any ATC simulation platform.

The Federal Aviation Administration (FAA) evaluates proposed new Air Traffic Control (ATC) technologies and procedures (henceforth referred to as ATC tools) that have the potential to enhance safety and efficiency in the United States' air transportation system. Evaluations that examine human performance and other aspects of human factors are important in that they produce evidence for benefits that a new ATC tool may provide along with potential problem areas that will need to be addressed prior to deployment. The FAA has a long history of using human-in-the-loop (HITL) simulations to evaluate aspects of ATC (Anderson & Vickers, 1953). Although the use of simulations has limitations and is not without its own challenges (Buckley, DeBaryshem, Hitchner, & Kohn, 1983), it remains the primary means of evaluating ATC tools prior to using them with live air traffic and can provide cost savings when conducted prior to the completion of costly prototypes.

Additionally, new ATC tools should be evaluated against situations as close as possible to those that are likely to occur in real air traffic operations. Events that perturb the status quo but would not be considered rare in the context of ATC operations, such as convective weather, medical emergencies, equipment malfunction and air traffic compression, are sometimes referred to as off-nominal events (Burian, 2008). Researchers conducting HITL evaluations of new ATC tools should use scenarios containing off-nominal events, in addition to other scenarios that do not include off-nominal events, whenever possible.

Proposals to change the ATC system may come from a variety of developers and researchers both within and outside the FAA. These organizations coordinate some aspects of their development and evaluation work formally with the FAA, and infrequently with each other, and coordinate other aspects only occasionally. There are a variety of ATC simulation platforms, including those commercially available and those developed in house by the organizations. Furthermore, there are multiple accepted ways to measure many human factors variables such as workload (Stein, 1985, and Hart & Staveland, 1988). The outcome of the situation in the ATC industry has been that the different industry groups that develop and propose new ATC tools usually do not have access to the same ATC Subject Matter Expert

(SME) participants, use the same scenarios or even the same airspace, may not use the same ATC simulation platforms, and may not collect and report the same performance measures as other industry groups when conducting evaluations. Attempting to compare results from HITL simulations that use different airspace and air traffic situations and report different types of performance measures presents an additional challenge to ATC tool evaluators and decision makers at the FAA, in addition to those inherent to the HITL simulation evaluation method.

This project provides organizations, both within and outside the FAA, with a standard set of scenarios against which new ATC tools can be evaluated. If organizations that run HITL simulations to evaluate their proposed ATC tools use these scenarios, it will facilitate comparisons between various proposed tools going forward.

Methods

This section describes the methods we used to identify airspace suitable for evaluating new ATC tools, the traffic volume and pattern for the scenarios, the number and type of scenarios created, and the off-nominal events included in the scenarios.

Airspace Selection

The first step in the creation of scenarios was the selection of airspace in which the scenarios would take place. The use of generic (designed by researchers) airspace was considered as a possibility as it has certain advantages. Generic airspace allows researchers to build made-to-order challenges into sectors and, given that no controller would have encountered the airspace outside of a lab, controller participants would all have the same level of unfamiliarity with the sectors. However, we ultimately elected to use real world sectors. There is a finite number of sectors in the National Airspace System (NAS). It was decided that the benefits of allowing proposed ATC tools to be tested for their ability to solve real world air traffic issues and the face validity that accompanies the use of a real world sector would, for this project, outweigh the advantages of using a generic sector. But the benefits of simulating real world airspace could only be achieved if the airspace selected would provide enough real world challenges or opportunities to solve real air traffic issues. Therefore, we attempted to identify the busiest and most complex sectors in the NAS.

Finding a suitable Terminal Radar Approach Control (TRACON) airspace was fairly straightforward. We used the Air Traffic Activity Data System (ATADS) to identify the airport with the greatest number of operations annually for the year 2014. It follows that this facility's TRACON would also be the busiest. The airport identified was Hartsfield–Jackson Atlanta International Airport (ATL). We selected the TRACON arrival airspace due to the preponderance of tools proposed for this type of operation.

We attempted to identify complex en route sectors by contracting an analysis of sectors in the three busiest Air Route Traffic Control Centers (ARTCCs): Atlanta (ZTL), Chicago (ZAU), and New York (ZNY). The sector analysis examined air traffic characteristics across a two-year time span (2013 to 2015). Traffic characteristics considered included Average Number of Aircraft in the sector per hour, Number of Climbing or Descending Aircraft per hour, Number of Potential Aircraft Conflicts per hour, and Number of Adjacent Sectors with which that sector controller would have to coordinate. Sectors with an average of fewer than 25 aircraft per hour were eliminated as possible candidates for simulation because of insufficient activity. The remaining sectors were compared with regard to how many climbing and descending aircraft and how many potential conflicts occurred per hour. We decided that, to increase opportunities to evaluate a wide variety of en route ATC tools, it would be necessary to provide scenarios for both a low altitude and a high altitude en route sector. Low and high altitude sectors have different characteristics that may differentially affect the way new tools are used or may differentially affect their utility for resolving the problem the tools were created to resolve. Certain sectors in both ZAU and ZNY were comparable in complexity, depending upon how one weighted the selected traffic characteristics. In ZNY, however, a low altitude sector and a high altitude sector on the candidate list were adjacent to each other. Since the two sectors were adjacent to each other, selecting them would create the possibility of simulating traffic through the two sectors simultaneously and, thus, provide an opportunity to collect data regarding coordination between sectors. The candidate ZNY sectors adjacent to each other were presented to project sponsors at the FAA who concurred with their selection.

Defining Scenarios and Scenario Events

Events are occurrences of interest scripted to take place during a scenario. We began the identification of suitable off-nominal events by using an event list collected during a previous project (Crutchfield & Pfleiderer, 2009). This list was created from the input of controller, pilot, and weather SMEs across five knowledge elicitation sessions that occurred during 2008. We updated that list using a hazard analysis of new ATC tools associated with NextGen (Sawyer, Berry, & Blanding, 2010). As our scenarios are meant to be used during HITL simulations, any event that specified a scripted error on the part of controller participants was removed from the list although pilot errors or errors on the part of controllers for adjacent scenarios were retained. Other events that we dropped from the list were events which we deemed to occur too rarely to be considered off-nominal (e.g., special handling of Air Force 1) or that would result in such a significant change to operations that the situation might be considered a better measure of emergency procedures than of a new ATC tool for normal operations (e.g. aircraft hijacking). Some of the events from the 2008 list required highly similar responses from controllers encountering them. In these cases a single representative event was selected from the group of similar events.

It is not likely that ATC tool evaluators will have the resources necessary to see how well a new tool performs during all of the off-nominal events identified. Furthermore, it was beyond the scope of our project to provide the number of scenarios necessary to cover all of these events. Therefore, we decided to select three high profile off-nominal events that should be included, along with a time period with no off-nominal events, in the standardized scenarios. The scenario that includes a time with no off-nominal events allows the ATC tools to demonstrate the benefits they can provide under ideal traffic conditions. The off-nominal events selected were Pop-up Storm, High Traffic Load, and Equipment Failure. Severe weather occurs somewhere in the NAS on a frequent basis and has the potential to impact traffic flows across the NAS for many hours. Additionally, evaluators and decision makers are interested in knowing how new ATC tools will perform in the face of high air traffic loads predicted to occur years into the future. Concerns about how the ATC system recovers during an equipment failure make it important to include a failure-related event as well.

We determined that all scenarios should be designed to be 40 minutes in length to minimize the amount of time controller participants are needed, while providing sufficient time to collect a useful amount of performance data. We determined that two versions of each evaluation scenario should be developed for each sector so that one version could be used as a baseline condition while the other could be used with the new ATC tool(s).

Air Traffic

The scenarios developed for this study were created by a retired controller employed with the ATAC Corporation using an I-Sim simulator provided by Kongsberg Geospatial. This retired controller had no experience controlling traffic in any of the three selected airspaces. The SME was directed to develop 40-minute long scenarios from real world air traffic data recorded in the summer of 2014 for the specified sectors using Performance Data Analysis and Reporting System (PDARS). Once the draft scenarios were created, the SME used WebEx to run them for other retired controllers to review. During this review, a retired controller from ZNY, familiar with the two selected sectors, reviewed the corresponding en route scenarios and a retired controller from our selected Atlanta TRACON approach control sector reviewed the TRACON scenarios. While they watched the scenarios, the retired controllers noted which flights they believed needed to be changed in some way to achieve the desired degree of realism in the scenario. The SME who developed the scenarios made changes to the scenarios in response.

Next, we used a second set of retired controllers familiar with the respective airspace and sectors to run the scenarios again. These scenario runs were conducted using a high fidelity simulation of an En Route Automation Modernization (ERAM) workstation and of a TRACON (Standard Terminal Automation Replacement System (STARS) workstation, again provided by Kongsberg Geospatial. The scenario runs used live pseudo-pilots to perform the associated flight deck/controller communications. Comments on how to improve these scenarios were collected from this second set of retired controllers and the scenarios were changed accordingly.

Lastly, we used a third set of retired controllers, one experienced with the ZNY sectors and another experienced with the selected Atlanta TRACON airspace, to run the scenarios with pseudo-pilots and make comments. We used these comments to make any final adjustments.

The second and third set of retired controllers were also asked to help us create presentations to be used in familiarizing controllers naïve to the selected airspaces with Letters of Agreement (LOAs), traffic flows, sector boundaries, and other types of information necessary to be able to control the simulated traffic. The familiarization material was then presented to retired controllers naïve to those airspaces who subsequently controlled two scenarios from each airspace. These naïve controllers were interviewed afterwards to identify information in the presentations that needed further clarification or recommend additional information that controller participants would need to be able to successfully control traffic in these sectors.

Results

Airspace Materials

We identified airspace at Atlanta TRACON A80 and New York ARTCC ZNY10 (a high altitude sector) and ZNY27 (a low altitude sector) to represent in the standardized scenarios. We developed Microsoft Excel files that include the sector boundaries, altitude definition, waypoints and fixes, routes, airports, and winds for each airspace. We also developed materials to familiarize participants with the airspaces being represented.

Air Traffic Scenarios

We developed and validated six scenarios suitable for use in evaluating TRACON tools, six scenarios suitable for use in evaluating en route tools in low altitude airspace, and six scenarios suitable for use in evaluating en route tools in high altitude airspace.

Two moderate traffic load scenarios for each sector were scripted without any off-nominal events. Two scenarios for each sector represented a moderate traffic load with the addition of a severe weather system that impacts operations in the sectors. Two scenarios for each sector have a traffic load 15% higher than the average that occurred in 2014. This traffic level represents what is predicted to occur in the year 2025 (Federal Aviation Administration, 2016).

The first 20 minutes of the moderate scenarios without off-nominal events enable new ATC tools to demonstrate the benefits they can provide under ideal traffic conditions. Our intent is that evaluators add their own equipment failure event to the second 20 minutes of these two scenarios. It was not possible to predict all the types of tools that may be evaluated with these scenarios, and selection of an unrelated type of equipment failure event would result in a less meaningful evaluation. Therefore, we suggest that evaluators include their own customized equipment failure event directly related to the ATC tool being evaluated.

Additionally, two moderate traffic load scenarios for each sector were developed as examples of scenarios that can be used to familiarize controller participants who are naïve to a given sector with the sector operations and traffic flow and also to familiarize them with the new ATC tool(s) being evaluated. Evaluators are encouraged to create more familiarization scenarios given available time and resources.

Discussion

We intend the scenarios to be used by a variety of organizations, both within and outside of the FAA, when evaluating new ATC tools. In so doing, this will foster a greater opportunity to compare controller performance associated with a wide variety of new ATC tools. The scenarios and other materials provided by this project were designed so that evaluators could either use controller participants who are familiar with controlling traffic in the sectors represented without any additional training, or use other participants who are naïve to the sectors but who can learn about them through familiarization materials and training scenarios. When running the HITL scenarios, it is expected that the evaluators will use a repeated measures design where every participant runs every scenario in turn. It is expected that evaluators will run one of each type of scenario provided (moderate traffic, weather, busy traffic) as a baseline using current technologies and procedures and run a second scenario (moderate traffic, weather, busy traffic) in an experimental condition that includes the use of the new ATC tools. It is further expected that the order of the scenarios used (baseline vs. new ATC tool) will be counterbalanced across participants to further control for differences in difficulty level that may inadvertently exist in the scenarios.

Although our primary goal was to provide scenarios that would allow the evaluation of the effectiveness of new ATC tools under conditions that might stress them, another goal was to fashion the experimental scenarios to be independent of any new ATC tool being evaluated. In some cases, the change to be evaluated may have an impact on air traffic flows into the airspace or on the structure of airspace objects (such as routes) within the airspace being represented itself. In these cases, it is suggested that when running the baseline condition, evaluators use the airspace as provided. When running the condition that uses the new ATC tool, evaluators are justified in changing the sequencing or spacing of aircraft entering into and operating within the airspace or the routes and/or airspace objects in the airspace, if the changes are similar to changes that would be made to any airspace using the new ATC tool. The number, type, and destination of the aircraft should not be changed.

Comparisons of tools evaluated using these standardized scenarios would be facilitated if evaluators collect and report standardized performance measures as well. Preliminary work to identify appropriate performance measures was done as part of this project. Further work, however, is needed to provide evaluators with details required to assure the measures are fully comparable.

Acknowledgments

This research was supported and funded by the FAA NextGen Organization’s Human Factors Division, ANG-C1.

References

Anderson, C. & Vickers, L. (1953). *Application of simulation techniques in the study of terminal area air traffic control problems*. (Report No. 192). Washington, DC: Civil Aeronautics Administration Technical Division.

Burian, B.K. (2008). Perturbing the system: Emergency and off-nominal situations under NextGen. *International Journal of Applied Aviation Studies*, *8*(1), 114-127.

Buckley, E. P., DeBaryshe, B. D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time air traffic control system simulation* (DOT/FAA/CT-83/26). Atlantic City International Airport: Federal Aviation Administration Technical Center.

Crutchfield, J. M., & Pfleiderer, E. (2009). *CAMI Far-Term TBO Scenarios*. Oklahoma City, OK: Authors.

Federal Aviation Administration. (2016). *FAA Aerospace Forecast Fiscal Years 2016-2036*. Washington DC: Forecasts and Performance Analysis Division, APO-100. Retrieved from http://www.faa.gov/data_research/aviation/.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Hancock, Peter A.; Meshkati, Najmedin. *Human Mental Workload* (PDF). Advances in Psychology. 52. Amsterdam: North Holland. pp. 139–183.

Sawyer, M., Berry, K., & Blanding, R. (2010). *NextGen Human Hazard Assessment Report*. Washington DC: TASC Inc.

Stein, E.S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport: Federal Aviation Administration Technical Center.