10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference
13 - 15 September 2010, Fort Worth, Texas

AIAA 2010-9293

Comparison of Ground-Based and Airborne Function Allocation Concepts for NextGen Using Human-In-The-Loop Simulations

David J. Wing[*]
NASA Langley Research Center, Hampton, VA, 23681

Dr. Thomas Prevot[†]
San Jose State University/NASA Ames Research Center, Moffett Field, CA, 94035

Dr. Jennifer L. Murdoch[‡]
NASA Langley Research Center, Hampton, VA, 23681

Christopher D. Cabrall[§], Jeffrey R. Homola[§], Dr. Lynne H. Martin[§], Joey S. Mercer[§]
San Jose State University/NASA Ames Research Center, Moffett Field, CA, 94035

Sherwood T. Hoadley[**], Dr. Sara R. Wilson[††], Clay E. Hubbs[‡‡],
James P. Chamberlain[*], Ryan C. Chartrand[*], Maria C. Consiglio[*], and Michael T. Palmer[*]
NASA Langley Research Center, Hampton, VA, 23681

Investigation of function allocation for the Next Generation Air Transportation System is being conducted by the National Aeronautics and Space Administration (NASA). To provide insight on comparability of different function allocations for separation assurance, two human-in-the-loop simulation experiments were conducted on homogeneous airborne and ground-based approaches to four-dimensional trajectory-based operations, one referred to as 'ground-based automated separation assurance' (ground-based) and the other as 'airborne trajectory management with self-separation' (airborne). In the coordinated simulations at NASA's Ames and Langley Research Centers, controllers for the ground-based concept at Ames and pilots for the airborne concept at Langley managed the same traffic scenarios using the two different concepts. The common scenarios represented a significant increase in airspace demand over current operations. Using common independent variables, the simulations varied traffic density, scheduling constraints, and the timing of trajectory change events. Common metrics were collected to enable a comparison of relevant results. Where comparisons were possible, no substantial differences in performance or operator acceptability were observed. Mean schedule conformance and flight path deviation were considered adequate for both approaches. Conflict detection warning times and resolution times were mostly adequate, but certain conflict situations were detected too late to be resolved in a timely manner. This led to some situations in which safety was compromised and/or workload was rated as being unacceptable in both experiments. Operators acknowledged these issues in their responses and ratings but gave generally positive assessments of the respective concept and operations they experienced. Future studies will evaluate technical improvements and procedural enhancements to achieve the required level of safety and acceptability and will investigate the integration of airborne and ground-based capabilities within the same airspace to leverage the benefits of each concept.

---

[*] Research Engineer, Crew Systems and Aviation Operations Branch, NASA LaRC, MS 152, AIAA member

[†] Research Engineer, Human Systems Integration Division, SJSU/NASA ARC, MS 262-4, AIAA member

[‡] Research Psychologist, Crew Systems and Aviation Operations Branch, NASA LaRC, MS 152

[§] Research Associate, Human Systems Integration Division, SJSU/NASA ARC, MS 262-4, AIAA member

[**] Senior Aerospace Engineer, Dynamic Systems and Control Branch, NASA LaRC, MS 152, past AIAA Associate Fellow

[††] Research Engineer, Aeronautics Systems Engineering Branch, NASA LaRC, MS 238

[‡‡] Aviation Operations Specialist, National Institute of Aerospace, NASA LaRC, MS 152, AIAA member

American Institute of Aeronautics and Astronautics

Nomenclature

4D	=	Four Dimensional	JPDO	=	Joint Planning and Development Office
ADS-B	=	Automatic Dependent Surveillance Broadcast	LOS	=	Loss of Separation
AFR	=	Autonomous Flight Rules	M	=	Moderate duration scenario
ANOVA	=	Analysis of Variance	MACS	=	Multi Aircraft Control System
ANSP	=	Air Navigation Service Provider	MAP	=	Monitor Alert Parameter
AOL	=	Airspace Operations Laboratory	MCH	=	Modified Cooper Harper
AOP	=	Autonomous Operations Planner	MCP	=	Mode Control Panel
ASAS	=	Airborne Separation Assistance System	NASA	=	National Aeronautics and Space Administration
ASTOR	=	Aircraft Simulation for Traffic Operations Research	NextGen	=	Next Generation Air Transportation System
ATC	=	Air Traffic Control	nmi	=	Nautical miles
ATM	=	Air Traffic Management	RTA	=	Required Time of Arrival
ATOL	=	Air Traffic Operations Laboratory	S	=	Short duration scenario
ATOS	=	Airspace and Traffic Operations Simulation	STA	=	Scheduled Time of Arrival
			TCP	=	Trajectory Change Point
CARS	=	Controller Acceptance Rating Scale	TMX	=	Traffic Manager
ETA	=	Estimated Time of Arrival	TLX	=	Task Load Index
FAA	=	Federal Aviation Administration	TSAFE	=	Tactical Separation Assisted Flight Environment
FL	=	Flight Level			
FMS	=	Flight Management System	ZID	=	Indianapolis Center
ft	=	Feet	ZKC	=	Kansas City Center
HITL	=	Human in the Loop			

I. Introduction

THE Joint Planning and Development Office (JPDO) has identified the action area 'Air/Ground Function Allocation' as a high priority research need for the development of the Next Generation Air Transportation System (NextGen).[1] Function allocation is a system design question that seeks to determine how certain processes, roles, or responsibilities related to key functions of Air Traffic Management (ATM) might be redistributed among actors to sustain system performance at an adequate level, and perhaps enhance it substantially, as air traffic demand grows beyond the limits of the current ATM system. The air traffic growth expected to occur over the next two decades is significant. For the period 2010 to 2030, forecasters are projecting system capacity, measured in 'available seat miles' – the overall yardstick for how busy aviation is both domestically and internationally – to grow an average of 3.4 percent a year, or to nearly double in that 20-year time period.[2] Load factors are already high today, and thus most of this growth will be absorbed by additional aircraft operations, which equates to additional workload for air traffic controllers in today's operational environment. The substantial delays experienced in the summer of 2000 indicated an ATM system already nearing its capacity limits. As indicated by the trends of NextGen, a consensus has formed that the current methods of Air Traffic Control (ATC) and management, if left unmodified, cannot indefinitely sustain the projected traffic growth without inducing significant delays and inefficiencies.

Separating aircraft is the most important task for an air traffic controller in high density airspace, and it is one of the main factors in controller workload today. To keep the workload manageable in current operations, a Monitor Alert Parameter (MAP) for each sector of airspace is used as an operating capacity limit. A sector's MAP value reflects the maximum instantaneous aircraft count that can be safely handled by a sector controller. It is recognized that actual sector capacity is not defined by aircraft count alone, but also includes many additional factors that define the air traffic complexity in that airspace. Recent research was conducted to investigate this relationship between air traffic complexity and sector capacity.[3] In a Human-In-The-Loop (HITL) simulation, a complexity threshold for a simulated sector was determined by participant controllers in terms of their self-assessed workload as aircraft count was steadily increased. The complexity threshold was determined to be reached when the participant controller called for the aircraft insertion to be stopped. The experiment results showed that aircraft count can safely exceed the MAP value by several aircraft in low complexity scenarios but not when complexity was considered high. This

result indicates that, while some gains in capacity may yet be possible by better understanding and controlling the factors that govern traffic complexity, the approach of separating aircraft using current day techniques remains inherently limited by controller workload and will not be able to support the expected traffic growth.

Hence, separation assurance has become a significant field of research and development. Studying alternative methodologies to separation assurance equates to exploring a reallocation of separation-related functions between the principal actors, ground-based and airborne, as well as between humans and automation. To enable sustained growth, the JPDO is interested in such research that considers alternative allocations of the functions involved in separating aircraft in high density airspace. The National Aeronautics and Space Administration (NASA) is the lead agency tasked with pursuing the necessary research activities. Teams at two NASA research centers, Ames and Langley, have been studying advanced concepts for 'separation assurance' for many years and have developed two concepts to a fairly high level of simulation fidelity in their respective laboratories. The two concepts feature distinctly different allocations of separation-related functions between air and ground, as well as between human and automation. The two concepts will be described in further detail in the next section of the report. However, the following summaries illustrate the high-level similarities and differences between the concepts. Both concepts involve new automation capabilities and new procedures for the human participants, either controllers or pilots. The primary difference between the concepts lies in the location of these changes, in ground-based ATC facilities in one concept and distributed among aircraft in the other. An additional difference lies in the degree of automation autonomy; separation assurance is fully automated in the first concept with human handling of exceptions identified by the automation, and interactive in the second concept with human involvement in each trajectory change for separation.

In the concept *ground-based automated separation assurance*[4] ('ground-based concept'), ground-based automation and air traffic controllers manage the separation between all aircraft within a defined airspace. The automation system predicts aircraft trajectories, detects conflicts, computes resolution trajectories, and issues trajectory amendments automatically to aircraft by data link communication if the resolution is within acceptable limits. The air traffic controller monitors the operation and is available for providing services and handling exceptions, such as creating or approving trajectories when the initial automated solution is outside the predefined tolerances. For example, the automation may determine that a traffic conflict requires a 4000 feet (ft) altitude change, but has been configured to issue only altitude changes automatically if they are within 2200 ft. In this case, the controller can still approve the solution, or use the ground-based trajectory automation to look for a different trajectory change or move a different aircraft. The role of the pilot, unchanged from current day operations, is to execute the instructions in a timely manner. Automation systems onboard the aircraft are a Flight Management System (FMS) and an auto-flight system capable of flying trajectories that meet route, altitude and speed constraints received via data link from the ground-based automation.

In the concept *airborne trajectory management with self-separation*[5] ('airborne concept'), the pilot manages the separation for his or her aircraft supported by onboard Airborne Separation Assistance System (ASAS) automation. The function allocation between human and automation in the airborne concept differs slightly from the ground-based concept in that the pilot is involved in each conflict resolution, whereas conflicts in the ground-based concept are often resolved automatically without controller involvement. Using airborne surveillance information, the ASAS automation predicts aircraft trajectories, detects conflicts, alerts the pilot appropriately, computes resolution trajectory alternatives, and displays these alternatives to the pilot. The pilot selects from among the alternative resolutions and executes the new trajectory usually through the FMS and auto-flight system. On the ground side, the controller performs no separation function for the ASAS-equipped aircraft. The Air Navigation Service Provider (ANSP) supplies ASAS-equipped aircraft with any trajectory constraints that may exist for traffic flow management purposes. The most common constraint would be a Required Time of Arrival (RTA) at an airspace boundary where ground-based control of the aircraft will resume.

Research of these concepts over the years has involved a number of separate HITL simulations of each concept. The simulation facilities used in these studies focused their highest modeling fidelity on the aspects most significantly changed for each concept, i.e., on the ground-based systems for the ground-based concept and the airborne systems for the airborne concept. Targeting fidelity in this way enabled detailed evaluations of automation tools and procedures by certified professional controllers and active commercial transport pilots using interfaces similar to their operational equipment. Consequently, the concepts have reached their respective levels of maturity in simulation facilities having different emphases, which becomes important when comparing the performance of the concepts with each other.

Such a comparison of concepts is desired by the JPDO to address the *"lack of clarity in the allocation of new functions to the aircraft and flight crew (includes human/automation as well as avionics/ground automation allocations)."*[1] NASA's ATM research emphasizes a range of airborne and ground-based capabilities. The user

community of NextGen will be as diverse as it is today, if not more so. The variations in operational models and business cases will dictate that systems appropriate for one user may not be suitable for another. This diversity calls for a range of available operational modes, within which each user can select the operation that best meets their needs. NASA's research seeks not to compete one concept against the other for an eventual victor, but rather to illuminate the characteristics of each concept through simulation and experimentation and to mature their designs to eventually enable multiple viable modes of operation for NextGen including mixed operations.

To support this need, NASA is pursuing a series of coordinated simulations to compare and integrate airborne and ground-based concepts. Two initial HITL experiments, a "ground-based experiment" and an "airborne experiment," have been jointly designed and conducted to make comparisons where possible and identify where they cannot. These experiments simulated nominal, homogeneous operations in which all aircraft within each experiment were separated in the same ground-based or airborne manner. The principal goal of these initial experiments was to determine how to achieve comparability of results from different simulation platforms while providing baseline results for future experiments. Plans for future studies include introducing mixed operations and off-nominal conditions. The human operators that served as subject participants in these initial experiments were controllers for the ground-based concept and pilots for the airborne concept. Data were collected to assess the acceptability of the concepts to the human operators and to evaluate the effectiveness of the associated technologies and procedures in high density airspace.

This paper provides an overview of the joint experiment design, a discussion of comparability, and comparable as well as non-comparable data. Section II presents the two operational concepts in greater detail. Section III describes the simulation facilities and automation tools. Section IV describes the common and differing aspects of experiment design. Section V presents detailed results and discussion. Section VI presents a summary of findings, and Section VII presents conclusions and recommendations for future function allocation research.

II. Operational Concepts

The two operational concepts are described as they were presented to the experiment subjects (i.e., participant pilots or controllers). Following this description, conceptual issues related to comparability of experimental results are discussed. It should be noted that the concept descriptions herein are generally limited to the aspects represented in these experiments and do not comprehensively address all aspects of an operational concept. As these experiments were designed to simulate homogeneous operations in nominal conditions, many details involving mixed operations and off-nominal conditions were not included in the experiment or the concept descriptions below.

A. Ground-based Concept
Ground-based automated separation assurance is a concept that involves a centralized system with ground-side automation components that monitor and/or manage nominal trajectory-based operations of equipped aircraft, while the controller handles off-nominal operations, provides additional services, and makes decisions on situations that are presented to him/her.[4] The separation responsibility resides with the ANSP, here meaning both the air traffic controller and the ground-based automation. The primary difference to today's system is that the ground-based automation is responsible for conflict detection, and separation assurance automation generates conflict resolution trajectories integrated with data link. The modified trajectories are sent to aircraft either by the controller or directly by the ground-based automation, whenever certain predefined criteria are met. The flight crews' responsibilities related to separation assurance do not change from the current day.

1. Enabling Environment
The concept of automated separation assurance is enabled by integrating controller workstations, ground-based automation, data link, flight management automation, and flight deck interfaces. The ground automation creates, maintains, and communicates trajectories for each flight. The air traffic environment is generally based upon the mid-term environment for the high altitude airspace outlined by the Federal Aviation Administration (FAA)[6] and assumes the following characteristics: Each aircraft entering high altitude airspace is equipped with an FMS that meet a required navigation performance value of 1.0 and have integrated data link for route modifications, frequency changes, cruise altitudes, climb, cruise, and descent speeds, similar to those available in current day Future Air Navigation System technology. Data link is the primary means of communication, and all aircraft are cleared to proceed, climb, cruise and descend via their nominal or uplinked trajectories. High accuracy surveillance information for position and speed is provided via Automatic Dependent Surveillance Broadcast (ADS-B) or a comparable source. In order to reduce trajectory uncertainties, FMS values for climb, cruise/ descent speeds, and weight are communicated to the ATC system. The goal is to make conflict detection highly reliable and to detect

trajectory-based conflicts with enough time before initial Loss Of Separation (LOS). However, some sources of trajectory uncertainties remain and include flight technical differences, trajectory mismatches between the air and the ground, inaccurate performance estimates and inaccurate weather forecasts used by the air and the ground automation. A conformance monitoring function detects off-trajectory operations and triggers an off-trajectory conflict probe. The trajectory generation function used for conflict resolution and all trajectory planning provides FMS compatible and loadable trajectories. These trajectories account for the nominal transmission and execution delays associated with data link messaging. Automated trajectory-based conflict resolutions are generated for conflicts with more than three minutes to initial loss of separation. When conflicts are detected within a short time before LOS, an automated tactical conflict avoidance function can generate heading changes and send this information to the flight deck via a separate high-priority data link path.

2. Roles and Responsibilities

The ANSP is responsible for maintaining safe separation between aircraft. The ground automation is responsible for detecting "strategic" medium-term conflicts (typically up to 15 minutes) between all trajectories and for monitoring the compliance status of all aircraft relative to their reference trajectory. The ground automation is also responsible for detecting "tactical" short-term conflicts (typically 0 to 3 minutes) between all aircraft. Whenever the ground automation cannot resolve a conflict without controller involvement, it must alert the controller early enough so that he or she can make an informed decision and keep the aircraft safely separated. The ground automation is responsible for alerting controllers to problems and exceptional situations.

The controller is supervising the automation and is responsible for making decisions on all situations that are presented to him or her by the automation, flight crews or other ANSP operators, such as controllers or traffic managers. Additionally, they are responsible for providing service in time-based metering and weather avoidance operations. Issuing control instructions to non data-link-equipped aircraft is also the responsibility of the controller. The controller can use conflict detection and resolution automation to generate new trajectories for any aircraft. Controllers use data link to communicate with equipped aircraft and voice for non data link-equipped aircraft.

Flight crews are responsible for following their uplinked (or initially preferred) trajectory within defined tolerances and for the safe conduct of their flight (just like today). Flight crews can downlink trajectory change requests at any time. The ground automation probes the request for conflicts without involving the controller. If the requested trajectory is conflict free, the automation uplinks an approval message, otherwise it alerts the controller that there is a trajectory request to be reviewed.

3. Air Traffic Controller Workstation

Figure 1 depicts the air traffic controller workstation prototype designed for the above distribution of roles and responsibilities. Aircraft that are managed by the automation within the controller's sector are displayed brighter than the aircraft outside that area, which are low-lighted. Additional information in data tags and colors are used to draw the controller's attention to a specific problem. The display is designed for general situation awareness and management by exception. The sector displayed in the figure contains approximately three times as many aircraft as can be controlled within this sector in current day operations.

All functions for conflict detection and resolution, trajectory planning and routine operations are directly accessible from the controller display. Transfer of control and communication between sectors is conducted by the automation close to the sector boundaries. Nominally, aircraft are
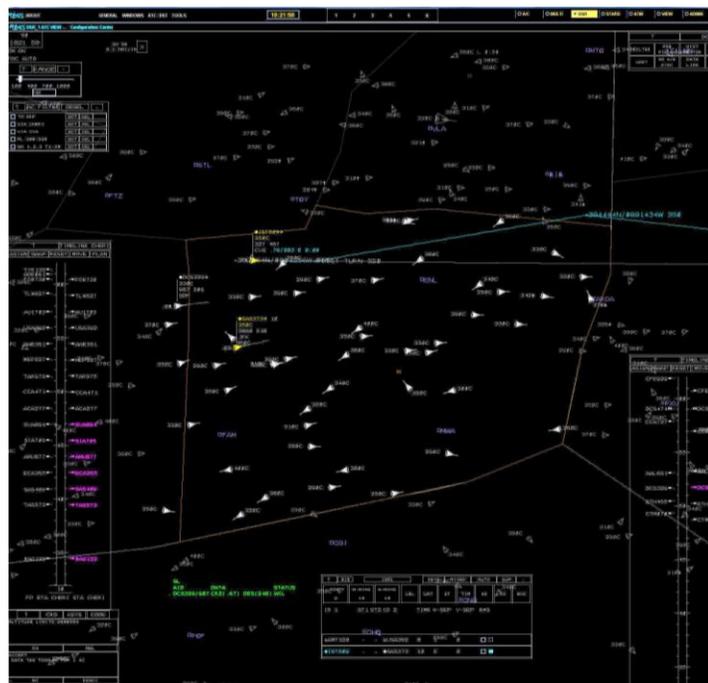
Figure 1. Controller display with conflict list, timelines and a provisional trajectory.

5

American Institute of Aeronautics and Astronautics

displayed as chevrons with altitudes, a design originally developed for cockpit displays of traffic information[7]. Traffic conflict information, hazard penetration and metering information is presented where applicable. Full data tags are only displayed in short-term conflict situations, or when the controller selects them manually. Time-based metering is supported via timelines and meter lists. The timelines show aircrafts' estimated and scheduled arrival times at specific fixes, which are often meter fixes into congested airports.

The controller can request trajectories to avoid traffic conflicts and weather hazards and to solve metering conflicts via various easy-to use mechanisms, using keyboard entries, data tag items, the conflict list, or the timeline. The automated trajectory-based conflict resolutions are generated by an autoresolver module originally developed as part of the Advanced Airspace Concept.[8] When initiated by the controller, the automatically generated trajectory becomes a provisional trial-plan trajectory (e.g. the cyan line in Fig. 1). The controller can then modify and/or uplink the trajectory constraints to the aircraft. All trajectory changes are immediately probed for conflicts and provide real-time feedback on their status, before they are sent. Therefore, the tools are designed to be interactively used.

B. Airborne Concept

Airborne trajectory management with self-separation is a concept in which the pilots of equipped aircraft actively manage their trajectories while maintaining separation from traffic. Such aircraft are said to be operating under 'Autonomous Flight Rules' (AFR). The AFR pilot is responsible for separation but relies on flight deck automation for critical support functions such as monitoring the traffic, detecting conflicts, computing resolution alternatives, and probing for potential new conflicts in pilot-proposed maneuvers. The controller does not have any role in separating AFR aircraft. However, the concept assumes that trajectories are subject to ground-based flow management constraints that enable orderly transitions to and from ground-managed terminal airspace. Because no single person or automation system manages separation for all aircraft within an entire airspace region, the concept is a distributed approach to separation. Each aircraft manages its own separation, and automated methods of implicit coordination are implemented to ensure the actions of multiple aircraft are complementary. Thus, the primary differences from today's operation are the highly distributed allocation of the separation function to equipped aircraft and the human's reliance on automation for performing key support functions.

Although the concept of self-separation is vehicle-centric in many ways, inter-aircraft coordination and system optimization also play central roles. Coordination is applied implicitly through, for example, right-of-way rules built into the pilot's automation tool. Right-of-way is typically based on conflict geometry (similar to Visual Flight Rules) but may also be designed to incorporate a variety of additional system-level optimization objectives, such as giving arrival traffic priority over crossing traffic. Since the right-of-way rules are encoded directly in the automation and do not rely on human memory or interpretation, the rule set may be as complex and extensive as needed to provide the desired system-level balance of equity and efficiency.

AFR operations may occur in homogeneous airspace (all aircraft as AFR) or in mixed-operations airspace (AFR and ground-managed aircraft sharing the airspace without segregation). Mixed operations are enabled by rules and coordination mechanisms between the AFR aircraft and the ANSP, and these issues will be the focus of the next function allocation joint simulations.

1. Enabling Environment

AFR operations are enabled first and foremost by airborne surveillance. The recent mandate by the FAA that all aircraft operating in transponder airspace be equipped with ADS-B Out by 2020 produces a surveillance-rich environment for aircraft equipped with ADS-B In, and the airborne concept envisions making full operational use of this information. Each AFR aircraft receives rapidly updated state data broadcasted by all aircraft within reception range, i.e. the same position, altitude, groundspeed and vertical speed information used by the broadcasting aircraft for guidance, navigation, and control. In addition to this state vector information, the airborne concept also makes use of four-dimensional (4D) trajectory information sent as a sequence of Trajectory Change Points (TCP). The AFR tool uses TCPs to reconstruct the near-term (~25 minutes) trajectories of broadcasting aircraft for use in conflict detection and resolution support. TCP broadcast is not currently in the ADS-B Out mandate, but it is included in ADS-B standards as a future capability and is expected to result in better system performance.

AFR operations would occur primarily in the en-route phase of flight but could also include segments of the departure and arrival phases. With the arrival phase comes the requirement for AFR aircraft to sequence into terminal airspace, and in high density environments, AFR aircraft would transition to ground-based control in the vicinity of a metering fix. A typical trajectory constraint for AFR aircraft would be an RTA at the metering fix, delivered from the ANSP to the aircraft through data link prior to the top-of-descent.

2. Roles and Responsibilities

The pilot's responsibility is defined by four AFR procedural requirements or "rules." These rules were presented to the pilot participants in this experiment in the following priority order.

AFR Rule #1: The pilot must resolve traffic conflicts without delay when notified by the AFR tool. The pilot may select from a set of tool-provided resolutions and must execute the selected resolution in a timely fashion. Figure 2 shows the aircraft navigation display with an example of a conflict (white segment and aircraft) and a resolution (blue line). Resolutions from the AFR tool are nominally in the form of a strategic, closed-path route that is uploadable to the FMS, but may also be tactical, open-form instructions, akin to controller vectoring, to be executed using Mode Control Panel (MCP) guidance.

AFR Rule #2: The pilot must use the AFR tool to check any trajectory changes for conflicts before executing the change (for turbulence, weather hazards, fuel efficiency, etc.). Specifically, this rule prohibits a trajectory change that would create a so-called 'Level 2' conflict alert for any aircraft, including their own. This notification level indicates greater urgency to resolve because of reduced time until LOS. Less urgent 'Level 1' conflicts may be temporarily created but must then be resolved according to AFR Rule #1.

AFR Rule #3: The pilot ensures that the aircraft's trajectory will conform to ATC constraints, should any be in effect. ATC constraints may include special use airspace restrictions or RTAs for arrival metering. Any changes to the trajectory must ultimately support RTA conformance, otherwise ATC must be notified if the constraint becomes unachievable. If a path stretch is necessary to absorb a delay, the pilot must use the AFR tool to devise a conflict-free path, in accordance with AFR Rule #2.

AFR Rule #4: The pilot has the auto-flight system FMS-coupled as much as possible. This lowest priority rule, perhaps better described as a guideline, is intended to maximize the proportion of aircraft on a 4D trajectory, i.e., 'strategic' flight, while providing the pilot with the flexibility to operate in 'tactical' flight when necessary, i.e., using just MCP guidance. Aircraft flown using FMS guidance are anticipated to have more predictable and stable trajectories, thereby benefitting all operators and controllers using that airspace. AFR pilots are therefore encouraged to fly with the FMS engaged to the greatest extent reasonably achievable, and right-of-way preference will be given over those in a tactical flight mode.

The role of the ANSP in homogeneous AFR operations is primarily that of a strategic airspace resources manager. This role comes into play where contention exists for limited-capacity airspace resources, most commonly the arrival to busy terminal airspace. The ANSP manages this flow by establishing a meter list and communicating RTAs to AFR aircraft. Aircraft unable to meet their RTA are resequenced or redirected to a different arrival fix. The controller does not have any role in separating AFR aircraft.

3. Flight Deck AFR Tool

The AFR pilot depends on significant support from a suite of automation capabilities integrated into the aircraft's avionics system, i.e., the AFR tool. Each of the AFR rules described above requires supporting automation functionality. In general terms, the AFR tool maintains active predictions of the ownship trajectory and those of all traffic aircraft within reception range, and it continually probes for traffic-to-ownship conflicts. For each detected conflict, right-of-way rules are applied to determine the burdened aircraft which in turn determines which aircraft is alerted first. At the appropriate time, the tool alerts the pilot to the conflict through displays (Fig. 2) and audible alerts. It also actively interprets flight crew inputs to the auto-flight system indicating a crew-proposed trajectory change or maneuver, and it indicates any resulting "provisional" conflicts prior to execution. For active conflicts, the tool produces resolution alternatives in the form of strategic (FMS) and/or tactical (MCP) actions for pilot review and selection. When time to LOS is short, the tool will favor tactical resolutions over strategic resolutions to ensure the flight crew has sufficient time to implement the resolution. In the process, any deviations from ATC constraints in effect are minimized, with the

Figure 2. Navigation display showing a conflict (white) and a resolution (blue).

priority given to the safety objective of maintaining separation. Additional tool functionality supports the pilot in achieving constraint conformance and in reestablishing a conflict-free strategic route from tactical flight. These capabilities will be described in more detail in Section III-B.

C. Comparative Discussion of the Operational Concepts

In contrast to current-day operations, the two operational concepts investigated in this function allocation study bear a mutual similarity of substantially enhancing the role of automation in separating aircraft. Although the airborne and ground-based automation system functions have many similarities, their designs are necessarily different for several reasons. For example, their information sources differ. The ground automation has uniform quality information of all aircraft in its domain. Its source for state information is ADS-B, the same as for the airborne automation. However, since it does not have access to the real-time performance data in the FMS of each aircraft and does not use TCP broadcast information to reconstruct trajectories, it generates trajectories based on estimates of aircraft performance and on previous trajectory amendments sent to the aircraft. Since the ground automation initiates or approves all trajectory amendments, it includes active and pending trajectories into its processing, whereas the airborne system processes only currently active trajectories for all traffic aircraft.

The airborne automation system, through access to its own aircraft's avionics systems, has certain aircraft-specific information not always available to ground systems. Onboard information, such as airframe and engine performance data, current weight, auto-flight settings, and measured atmospheric data, as well as information from ground systems on predicted winds, is used to predict trajectories with the highest possible accuracy. For traffic trajectory predictions, the automation reconstructs trajectories from TCPs generated by the avionics systems of the traffic aircraft (also with high accuracy). However, traffic trajectory predictions have reduced fidelity than ownship trajectory predictions because of assumptions made about the other aircraft's flight in between the broadcast TCPs. Additionally, the ground automation uses primarily the planned trajectory, which describes the trajectory the aircraft will fly if the flight crew makes all the inputs necessary to follow this trajectory, whereas the airborne system uses the so-called command-trajectory, which describes the trajectory the aircraft will fly if no further pilot inputs are made. These differences in information source can drive design differences between the airborne and ground-based automation systems which must be considered when comparing their performance.

Another difference to be considered between the concepts is the human role in the separation function. In both concepts, conflict detection is solely performed by the automation system, and neither the controllers in the ground-based concept nor the pilots in the airborne concept are responsible for detecting conflicts, nor are they liable for conflicts not detected or detected too late for resolution. Indeed, this point of similarity places a burden on automation systems development and distinguishes these two concepts from current day operations more than perhaps any other. From here, however, a difference emerges between the concepts in the human role. In the ground-based concept, most conflicts are fully resolved by the automation without knowledge or review by a human (the controller). The controller becomes involved if there is an exception presented to him or her by the automation, but these events are expected to be infrequent in a fully-equipped environment. The pilots in the ground-based concept would of course be aware of the trajectory change but would have no direct role or responsibility in the separation function (just like today). In the airborne concept, every conflict which the automation detects and presents will have the attention and involvement of a human (the pilot of the aircraft requiring a trajectory change), although one human will not have awareness of every conflict as in today's operations. The pilot's involvement is generally to execute one of the resolutions provided by the automation. At this point, the pilots' procedural duties in both concepts merge, i.e., execute the conflict resolution trajectory (or select from a small set) provided by the automation.

III. Simulation Descriptions

The airborne and ground-based concept experiments for this function allocation research activity were conducted during March 2010 in separate laboratories. The laboratories and simulation platforms are described, followed by a discussion of the implications of using separate laboratories for comparative function allocation research.

A. Ground-based Concept Simulation: Airspace Operations Laboratory

The ground-based concept simulation was conducted in the Airspace Operations Laboratory (AOL) at the NASA Ames Research Center, in Moffett Field, California.[9] The AOL specializes in developing and evaluating advanced concepts and technologies for NextGen air traffic control and management. The AOL currently investigates operational concepts in the areas of airspace super density operations, dynamic airspace configuration, multi sector planning, and separation assurance/function allocation.[10] The AOL can currently be configured with workstations

for up to 10 air traffic controllers, 2 supervisors, 4 traffic management coordinators, 4 confederate controllers, and 10 simulation pilots. The AOL is currently expanding to increase this simulation capacity by approximately 70 percent to enable parallel simulations for different focus areas as well as integrated cross facility simulations.

1. Simulation Platform

The AOL uses the Multi Aircraft Control System (MACS) with its networking supplement Aeronautical Data link and Radar Simulator (ADRS) as its simulation and rapid prototyping software.[11] This software is developed and maintained by AOL engineers and distributed to partners in government, industry and academia. The existing suite of capabilities allows researchers to configure a wide range of air traffic environments, from accurately emulating current-day operations to simulating many of the operations envisioned for NextGen as well as the transitional stages in between. MACS provides high-fidelity display emulations for air traffic controllers and managers as well as user interfaces and displays for confederate pilots and flight crew participants, and experiment managers, analysts, and observers. Scenario and target generation capabilities are also built into MACS, which are used to generate and run traffic problems tailored to the specified research project. The integrated data collection system is used to collect the quantitative measures of interest at each operator station as well as overall traffic progression, including aircraft states, conflicts, and sector counts.

In order to provide the required automation support to the controller, a new NextGen ATC workstation prototype was developed based upon an emulation of the operational en-route controller system. The workstation provides access to key functions that support the operator in managing high traffic densities effectively. Figure 1 earlier in this paper shows the controller display as implemented in MACS that was used for this research.

For this study, the AOL was configured with two participant control rooms, each hosting four air traffic control sector positions and one supervisor position. Four high and low altitude confederate controllers managed traffic flows in and out of the test sectors, and 10 general aviation pilots operated the aircraft throughout the test airspace. During the study runs, the two control rooms were run independently in separate "worlds," each with their own confederate controllers and simulation pilots. Figure 3 shows part of one of the air traffic control rooms during the study. Radar associate controllers (D-sides) were not staffed; each workstation shown was for one sector worked by a single radar controller (R-side). Figure 3 also shows the overall traffic situation projected on the wall. This wall display is configured and controlled by the supervisor behind the controller position.

Figure 3. Air traffic control room in the NASA Ames Airspace Operations Laboratory.

2. Ground-Based Automation System for Separation

The ground side automation system prototype used in this simulation represents a synergy between Erzberger's work on ground-based automated separation assurance [8, 12] and the AOL's prior human-in-the-loop research on interactive NextGen air traffic control technologies.[13] Erzberger's Advanced Airspace Concept is theoretically designed to provide fully automated separation assurance and air traffic control operations. The technologies developed in the AOL with the help of many controller-in-the-loop simulations were designed as highly responsive semi-automated decision support tools. The resulting superset of these tools enables simulating a wide range of concepts for function allocation between the controller and the automation. As indicated before, this simulation used a cooperative allocation of functions between the ground automation and the controller.

The ground automation created flight plan-based trajectory predictions for all aircraft from their present position to the destination airport. A conformance monitoring function compared each aircraft's actual position and velocity vector to its flight plan trajectory. When vertical non-conformance was detected, the trajectory generation process used estimates of the current vertical rate and the aircraft supplied target altitude to generate the prediction. When an aircraft was in lateral non-conformance, or in other words "off-track," the automation used the aircraft's supplied target heading to generate a flight state-based trajectory prediction for the next five minutes. Off-track aircraft create an undesirable state because the system has no medium term trajectory prediction. These aircraft were indicated prominently to the controllers, highlighting a need for implementing a new trajectory for the aircraft. For trajectory-

based concepts, ideally all aircraft are in lateral and vertical conformance so that their active trajectories are highly predictable.

The ground-based automation system consisted of a layered separation management approach of strategic and tactical capabilities. All currently active trajectory predictions were tested within each conflict detection cycle as to whether a LOS was likely to occur within the next ten minutes. If a conflict was first detected between four and ten minutes to initial LOS, the system invoked the strategic "autoresolver"[12] to determine the best overall conflict resolution according to its built-in heuristics. The conflict resolution tried to avoid traffic and meet any potential time constraints. In this study, if a resolution did not change altitude more than 2200 feet, change heading more than 60 degrees, violate a time constraint, or if unconstrained, cause an overall flight delay of more than 90 seconds, the automation created and sent a data link message to the aircraft that included all parameters that needed to be loaded into the aircraft FMS to compute a trajectory that sufficiently matches the ground-based trajectory. The tolerances for issuing resolutions automatically were chosen during simulations with retired controllers that were conducted in preparation of the study. The ground-based database of flight plans and trajectories was immediately updated so that the next conflict probe cycle would no longer flag this conflict and future conflict resolutions could take the new trajectory into account.

If a conflict resolution fell outside these parameters, it was flagged to the controller for review. The controller then used semi-automated functions (including the autoresolver) to evaluate different strategic options or approve solutions that were outside the tolerances that would have allowed the automation to issue them without controller involvement. If a conflict was predicted to occur within less than three minutes, the Tactical Separation Assisted Flight Environment (TSAFE)[14] module was activated and computed tactical heading changes for one or both of the aircraft involved in the conflict. In the current study, the automation sent the heading change(s) at two minutes to predicted LOS automatically, and the controller had no means to intervene. Not providing a manual override was an experimental design decision rather than part of the operational concept.

While detection and resolution of traffic conflicts was basically automated in the study, other tasks had to be conducted by the controllers using automated aids. One controller task was to manually create and send trajectories to put off-track aircraft back on a known trajectory. Another manually initiated task was arrival management. The ground-based concept experiment ran a script to use the exact same Scheduled Times of Arrival (STA) as the RTAs used in the airborne concept experiment. In order to make sure aircraft adhered to this schedule, controllers could invoke a meet time function that would combine the autoresolver logic with a speed advisory function and compute a combination of route, altitude, and speed change that would achieve the desired STA on a conflict free path. All functions were implemented in MACS. The autoresolver and TSAFE source code modules developed by Erzberger and Heere were integrated into the MACS Java code and used the MACS built-in trajectory generation and conflict probing functions.

B. Airborne Concept Simulation: Air Traffic Operations Laboratory

The airborne concept simulation was conducted in the Air Traffic Operations Laboratory (ATOL) at the NASA Langley Research Center, in Hampton, Virginia. This facility specializes in simulation-based research and development of advanced operational concepts employing ADS-B, but it is suitable for a variety of traffic operations research. For conducting batch and HITL experiments, it has over 300 computers, including 12 desktop pilot workstations, four of which are shown in Fig. 4.

1. Simulation Platform

The ATOL computing network hosts the Airspace and Traffic Operations Simulation (ATOS) platform.[15] ATOS provides a medium fidelity setting for studying the interactions of aircraft in a realistic ADS-B environment. The simulation networks multiple individual pilot stations called Aircraft Simulation for Traffic Operations Research (ASTOR) and a background traffic generator, named Traffic Manager (TMX).[16] ATOS supports the exchange of industry standard ADS-B reports (i.e., state vector, mode status, air referenced velocity, target state, and trajectory change reports) with standard message content and broadcast frequency.[17] For this experiment, frequency interference modeling was disabled, transmission range

Figure 4. Pilot workstations in the NASA Langley Air Traffic Operations Laboratory.

was fixed at 120 nautical miles (nmi) radius, broadcast rate of key reports was established at 1 Hz, and trajectory intent was broadcast up to 12 trajectory change points (i.e., effectively full trajectory intent).

The traffic aircraft in the experiment scenarios were provided by 12 pilot-station ASTOR computers (controlled by subject pilots), 63 batch ASTOR computers (non-piloted), and up to 557 lower fidelity aircraft modeled by a single TMX computer (also non-piloted). The function of the hundreds of TMX aircraft was to create the required traffic density and to provide ADS-B message broadcasts of aircraft performing AFR procedures. For this purpose, TMX aircraft included capabilities for conflict detection and resolution; however, their separation performance was not the subject of the airborne concept experiment.

Each ASTOR computer simulated one aircraft at medium fidelity. ASTOR has realistic displays and controls representative of a Boeing 777 and a six degree-of-freedom flight performance model representative of a medium-sized twin-engine transport aircraft. For the functions required in this experiment, the auto-flight system and FMS were fully functional. Its digital avionics data bus emulates realistic internal and external data communications.[18] ASTOR supports trajectory uplink and auto-load, a capability in this experiment used for loading RTAs. ASTOR is controlled by subject pilots using a desktop computer mouse. When used in the non-piloted batch mode, a pilot model automatically operates the flight controls in real time according to standard AFR procedures.

Since all aircraft in this experiment were AFR, and no aircraft entered terminal airspace during the simulation runs, no interactive ATC simulation component was required to accomplish the experiment objectives. Though ATC would normally assign schedule constraints dynamically, RTAs were predetermined for this experiment and sent as scripted data link messages at scheduled times during the appropriate scenarios. Each scenario contained 75 aircraft with destination airports in the vicinity of the experiment airspace for which arrival scheduling was scripted. Of these 75 aircraft, 12 were ASTORs flown by the subject pilots, and the remaining 63 aircraft were ASTORs flown by the pilot model. For runs that included schedule constraints, these 75 aircraft received data link messages containing their RTA. The remaining several hundred aircraft were departures and overflights and were not subject to arrival scheduling constraints.

2. Flight Deck Automation System for Separation

ASTOR contains a research-prototype ASAS, an AFR tool for the pilot called the Autonomous Operations Planner (AOP).[19] The AOP supplied the self-separation automation functions necessary for the pilot to meet the aforementioned four AFR rules: (1) resolve conflicts when notified, (2) check and clear all trajectory changes for conflicts before executing, (3) conform to ATC constraints, and (4) remain FMS-coupled whenever possible. The method of automation support for each rule follows.

Supporting AFR Rule #1, the AOP automated the process of conflict detection and crew notification by making trajectory predictions of the ownship and all traffic aircraft within ADS-B reception range, and by probing these trajectories for predicted LOS. A nominal look-ahead horizon of 10 minutes was used for detection, and the trajectory predictions were refreshed at least every 10 seconds. Buffers were applied to all trajectory segments to encompass prediction errors and minimize missed alerts.[20] The AOP notified the pilot of conflicts using textual, aural, and graphical methods, and at a time and notification level appropriate to the ownship aircraft's right-of-way and the conflict's level of urgency. A staggered notification scheme was used to provide implicit coordination and minimize simultaneous resolution actions by both aircraft.

Also for AFR Rule #1, the AOP automated the process of computing acceptable resolutions to a conflict. Using a layered separation management approach similar to the ground-based concept, two methods were available to the pilot: strategic and tactical.[21, 22] A strategic resolution is a modification of the FMS route, and in most situations, both lateral and vertical alternatives are computed. For each of these alternatives, a pattern-based genetic algorithm determined the fuel-optimal solution, and the pilot could upload either option to the FMS for execution. Figure 2 shows a Level 1 traffic conflict notification and an AOP-computed lateral strategic resolution that is ready for upload and execution in the FMS. To execute a vertical resolution, the pilot must also reset the MCP altitude window appropriately for the new altitude. A tactical resolution is a maneuver activated by the pilot typically through MCP 'track select' or 'flight level change' commands. The AOP will typically offer both lateral and vertical alternatives, computed by sweeping through possible track, altitude, and vertical speed changes until a conflict-free maneuver is found. At five or four minutes until LOS for the give-way and priority aircraft, respectively, the AOP will enter 'tactical override' mode and offer only tactical resolutions, given the short time remaining to resolve the conflict.

To help pilots meet AFR Rule #2 (i.e., clear all trajectory changes before executing), the AOP probed 'provisional' or "what if…" trajectories or maneuvers prior to execution. 'Planning' conflict symbology was displayed that indicated whether the proposed change would create a conflict and at what notification level. With this capability, pilots could safely investigate both FMS strategic trajectory changes and MCP tactical maneuvers

before executing them. In addition, to increase situation awareness, yellow bands were displayed on the flight displays to indicate ranges of tracks and vertical speeds that should not be selected since their selection would result in a Level 2 conflict for the ownship or another aircraft. Figure 2 shows such a 'maneuver restriction band' on the compass rose of the navigation display.

AFR Rule #3 (i.e., conform to ATC constraints) was supported with a combination of FMS and AOP capabilities. The FMS evaluated any required waypoint constraints, such as an RTA, and notified the pilot if the aircraft could not meet a given constraint with speed changes alone, e.g. too much delay to absorb by simply slowing down. The AOP then provided the capability to 'resolve' this 'unable RTA' situation with path changes using the same strategic resolution capability used for resolving traffic conflicts. In this experiment, only lateral 'Resolve RTA' solutions were enabled.

The AOP supported the pilot with adherence to AFR Rule #4 (i.e., remain FMS-coupled whenever possible) by providing a 'strategic reconnect' capability. This function was to be used following an MCP tactical maneuver that took the aircraft off its FMS path or cruise altitude. The AOP would construct a nominal reconnect path and probe it for conflicts, providing as well the capability to resolve them. It would also provide the pilot with guidance regarding the MCP settings required to again become 'fully coupled' to FMS guidance.

C. Comparative Discussion of the Simulation Environments

Conducting function allocation research in separate laboratories has advantages and disadvantages. The primary benefit is that each concept may be tested at the highest available fidelity for that concept, an important consideration for HITL safety-related operations research. Testing at higher fidelity means that the controllers in the ground-based concept or the pilots in the airborne concept will each use realistic and fully functioning interfaces and automation functionality, which allows them to better assess the strengths and weaknesses a particular concept would have in an operational setting. Comparing ground-based and airborne concepts in a single simulation platform is expensive to accomplish and would typically result in either (1) lower fidelity but balanced representation of both concepts with many simplifying assumptions, or (2) unbalanced representation of the concepts, with one concept being significantly compromised relative to the other because the higher fidelity platforms have favored either the ground side or airborne side in their development. For this study, the ground-based concept was tested in the same environment in which it was matured through prior human-in-the-loop studies: a high fidelity control room with workstations equipped with advanced automation aids. The airborne concept was also tested in the same environment in which it was developed and matured: multiple pilot stations with higher fidelity flight-deck interfaces, avionics functionality, and aircraft modeling.

A disadvantage of testing in separate environments is the unavoidable dissimilarity in modeling, an inherent result of the independent development history of the simulation platforms. For example, differences existed between MACS and ATOS (specifically the AOP) in the time at which tactical conflict resolution overrode strategic conflict resolution. In the ground-based automation, TSAFE was used starting at three minutes prior to LOS. In the AOP, the tactical override capability was used starting at four or five minutes prior to LOS, depending on right-of-way. These timings were considered integral to the design of each automation system and therefore were not candidates for synchronization. In another example, different methods used to account for trajectory prediction uncertainty. MACS added buffers to the separation standards, whereas the AOP placed buffers around trajectory segments and probed for separation loss between the buffers. The two approaches are fundamentally different and could not be made equivalent for this study. However, sources of uncertainty (e.g. wind errors) were minimized where possible to mitigate the difference.

A successful effort was made to make the descent profile characteristics of the MACS aircraft similar to the aircraft that were used as piloted ASTORS in the Langley experiment. In order to match the arrival profiles of the piloted ASTORs, the MACS B757 performance model was adjusted until the aircraft in both simulations flew almost identical descent paths and achieved comparable flight times from their initial position to the arrival fix, thus making arrival time comparisons more meaningful.

Where other modeling assumptions differed, compensations were made where possible. An example is the surveillance modeling. MACS has a higher fidelity RADAR model than ATOS, including modeling of RADAR noise and alpha beta tracking that cause realistic delays in position updates and track angle changes. ATOS has a higher fidelity ADS-B model than MACS, including probabilistic range and interference effects. To promote commonality, both MACS and ATOS were set up to use only ADS-B for surveillance, and ATOS used no interference modeling and a fixed cutoff range of 120 nm. In another example, MACS capability to model dynamic weather exceeded the capability in ATOS, and therefore the decision was made to remove dynamic weather from the common test matrix. Thus, fidelity was sacrificed in two instances in the interest of commonality.

IV. Experiment Design Commonality

The principal goal of these coordinated experiments was to determine how to achieve comparability of results from different simulation platforms while providing baseline results for future experiments. To assist in the comparability of results, the two experiments were designed with as many features in common as possible. The common aspects are described, followed by a discussion of those aspects that necessarily differed.

A. Common Aspects
The aspects of the two experiment designs that were established in common are listed in Table 1.

**Table 1. Common aspects of experiment designs.**

Airspace location and dimensions	*Rectangular region over Indianapolis and Kansas City Centers defined by: Latitude N36.5 to N40.5; Longitude W86 to W94; Flight Level 290 to 400*
Traffic density definition	*1.0x = 18 aircraft per 10,000 nmi^2; applied to the entire experiment airspace*
Aircraft types	*Equivalent aircraft types (e.g. B757-200) per call sign*
Traffic scenarios	*Initial states and flight plans of all aircraft*
Environmental conditions	*Winds westerly at 30 to 50 knots; Small variations by altitude; No lateral or time variation; No wind forecast error*
Separation standards	*Lateral: 5 nmi; Vertical: 1000 ft*
Test matrices and scenario durations	*Moderate (M): 2x2 within-subjects design with 2 replicates; eight 30-minute runs Short (S): 3x1 within-subjects design with 2 replicates; six 15-minute runs*
Independent variables and levels	*M: Traffic density: 1.5x, 2.0x; Scheduling assignment: No RTA/STA, RTA/STA = ETA S: Trajectory change event timing: None, Dispersed, Synchronous (Fixed: Traffic density, 2.0x; Scheduling assignment, RTA/STA = ETA)*
Trajectory change event	*Uplink of RTA/STA change indicating an arrival delay to be absorbed with a trajectory change*

1. Airspace and Aircraft
The experiments modeled the same airspace region, which straddled the eastern portion of Kansas City Center (ZKC) and the western portion of Indianapolis Center (ZID). The domain of interest was high altitude airspace above Flight Level (FL) 290. Figure 5 shows the test airspace. The rectangle indicates the common test airspace. In the ground-based experiment, sectors ZKC90, ZKC98, ZID80 and ZID81 were actively managed by air traffic controllers.
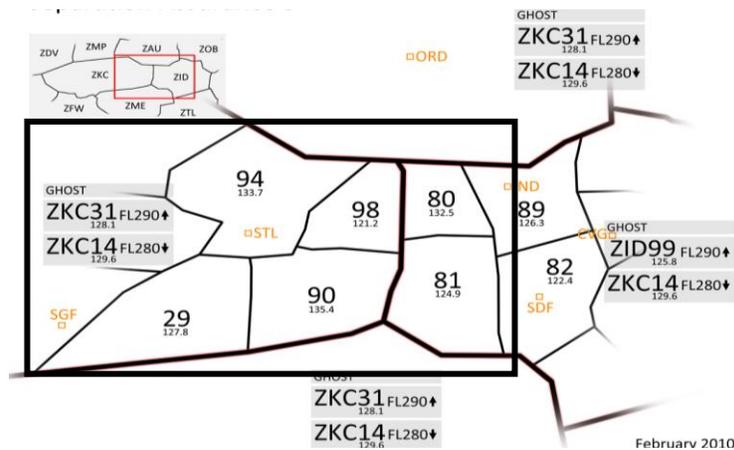

Figure 5. Common test airspace.
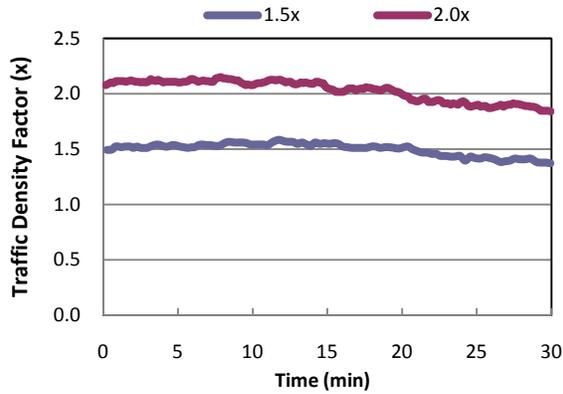
Figure 6. Traffic density sustained across the experiment airspace during the 1.5x and 2.0x scenarios.

Aircraft were permitted to climb to and descend from this region, but no aircraft cruised below FL290. The traffic scenarios were designed to approximate current day routing between city pairs and used current day waypoints. Traffic density was scaled up from current levels by proportionately increasing city-pair departure rates to reach the target aircraft counts equivalent to 1.5x and 2.0x current day maximum density. A common simplified definition was used for current day maximum traffic density, with 1.0x equal to 18 aircraft per 10,000

Figure 7. Aircraft count over time in the four ATC sectors during the 1.5x and 2.0x scenarios.

nmi[2*] or 164 aircraft for this airspace. The traffic density (i.e., aircraft count within the experiment airspace) was sustained for the duration of each simulation run by supplying new aircraft as other aircraft exited the airspace. Figure 6 shows the sustainment of traffic density within the experiment airspace, and Fig. 7 shows the traffic demand within each experiment sector. One item to note is that the experiment airspace encompassed a larger area or volume than the four test sectors. While traffic density of the entire experiment airspace was controlled to be relatively uniform, naturally occurring variations within and between the ATC sectors were observed, as shown in Fig. 7. Each ATC sector was therefore a unique environment, with some sectors having significantly more traffic than others. Similarly, each subject aircraft in the airborne experiment flew through a different part of the experiment airspace and was therefore exposed to a unique traffic environment ranging from light to heavy traffic density. Therefore, variation between results for each sector and each subject-aircraft are expected.

The aircraft in the two experiments' scenarios matched call sign by call sign, including aircraft type and initial conditions (i.e., initial position, altitude, speed, heading, and flight plan). Only the initial conditions can be guaranteed to match, since any trajectory or altitude change by any aircraft thereafter will affect how the remainder of the scenario unfolds. The two simulations were configured to use the same wind field, which was chosen to be laterally uniform, unchanging, and relatively mild. Although both simulations were capable of including differences between truth and predicted winds, wind error was not included to reduce testing time. The two simulations also assumed current day en-route separation standards were in effect.

2. Test Matrices and Scenario Definitions
The experiments shared two common test matrices, the "M" matrix, consisting of moderate duration (30 minute) scenarios, and the "S" matrix, consisting of short duration (15 minute) scenarios. The objective of the M matrix was to enable an assessment of nominal operations under different traffic loads and trajectory constraints. The test matrix, shown in Fig. 8, was a 2x2 within-subject design that varied traffic density (1.5x vs. 2.0x) and the inclusion of a scheduling assignment, i.e., an arrival time constraint (not included vs. included). The arrival time constraint was an STA in the ground-based experiment and an RTA in the airborne experiment set approximately equal to the aircraft's Estimated Time of Arrival (ETA) at the arrival metering fix. When included, the scheduling assignment was enacted in the airborne experiment by sending a data link message to the 75 aircraft arriving to seven common airports in the vicinity of the experiment region. In the ground-based experiment the scheduled times were entered
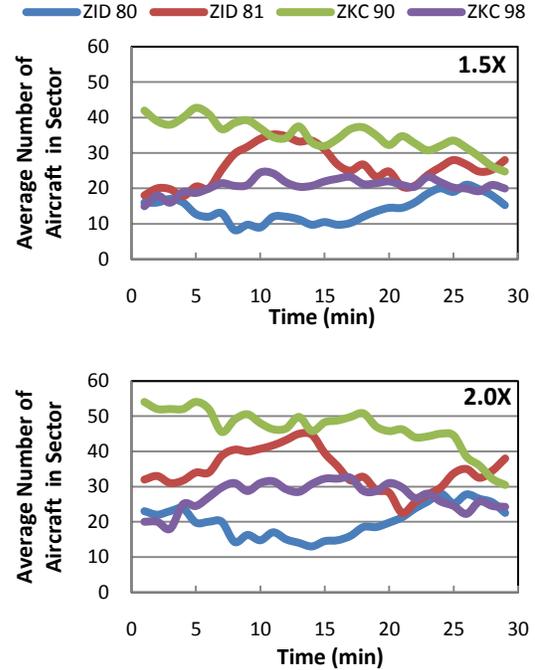
---

* In rough terms, this value was considered representative of a typical MAP value for a typical en-route sector size.

into the ground automation as meter times. Both experiments ran automated scripts to control the timing and exact values of these assignments. The 2x2 matrix contained two replicates for a total of eight scenario runs that each lasted 30 minutes.

The objective of the S matrix was to enable an assessment of the concept agility to large numbers of trajectory changes. The matrix, shown in Fig. 9, was a 3x1 within-subject design that varied a trajectory change event. The fixed conditions were a traffic density of 2.0x and the inclusion of an initial scheduling assignment approximately equal to the ETA.

The trajectory change event was a revision of the scheduling assignment representing an arrival delay of four to six minutes and was enacted in the same way as the scheduling assignments in the M scenarios. In addition to the baseline condition which excluded the event, the assignments were made at either dispersed times

Figure 8. Design matrix for the 30-minute M scenarios.

Figure 9. Design matrix for the 15-minute S scenarios. Traffic density was 2.0x in all three conditions of this matrix.

or synchronous times among the 75 affected aircraft. In the dispersed condition, no two aircraft received their arrival delay at the same time, whereas in the synchronous condition, all 75 aircraft received their new assignments at the same time. In the ground-based experiment, the new STAs were indicated in the controllers' arrival timelines, prompting them to initiate a trajectory change to absorb the delay. In the airborne experiment, the receipt of the arrival delay message by each aircraft triggered the process of generating a delay-absorbing maneuver (i.e., a path stretch) resulting in a trajectory change. For operational realism, only the timing of the initiating event was controlled, not the timing of the actual trajectory changes. Differences in the procedures and tools between the two concepts affected the timing of the actual trajectory changes relative to the initiating event.

B. Differing Aspects

Other aspects of the two experiment designs were necessarily different. These aspects are listed in Table 2 and discussed in the subsequent sections.

Table 2. Differing aspects of experiment design.

	Ground-based Concept Experiment	*Airborne Concept Experiment*
Subjects	*Controllers (sectors and supervisors)*	*Pilots (single operator, no crew)*
Subject quantity	*2 independent groups; 5 controllers per group; 1 group = 4 sector controllers, 1 area supervisor*	*4 independent groups; 12 pilots per group*
Subject total participation time	*2 groups simultaneously in 2 independent 'sector worlds'; single 8-day session*	*1 group per week for 4 weeks; each week contained one 3-day session*
Training time on concept, procedures	*3 days with 10 hours of hands-on training*	*1 day with 5 hours of hands-on training*
Aircraft monitored by subject participants	*4 airspace sectors (ZKC90, ZKC98, ZID80, ZID81) traversed by hundreds of aircraft per scenario; the remaining airspace was monitored by confederate controllers*	*12 aircraft per scenario; all remaining aircraft were automated and not piloted by humans*
Scenario difficulty and conflicts experienced	*Varied by <u>sector</u> assignment, between runs, and within runs*	*Varied by <u>aircraft</u> assignment, between runs, and within runs*

1. Subjects

Each HITL experiment recruited human subject participants most relevant for its concept: certified professional controllers for the ground-based concept and active commercial transport pilots for the airborne concept. Neither experiment involved both subject controllers and subject pilots.

In the ground-based experiment, trajectory amendments sent from the ground automation to aircraft were automatically executed; therefore no subject pilots were required for the experiment. Aircraft receiving tactical instructions did require human execution, but these actions were accomplished by non-subject general aviation pseudo-pilots. Pseudo-pilots were not considered a source of experimental data. Ten pseudo-pilots were present, each responsible for controlling between 10 and 70 aircraft within the test airspace and several hundred in the surrounding airspace. In addition, eight retired controllers served as confederates, controlling the airspace surrounding the subject-controller sectors. The confederate controllers were also not considered a source of experimental data.

In the airborne experiment, the only required ANSP (i.e., groundside) role involved the assignment and transmission of RTA constraints to the aircraft. Since these constraints were predetermined and fixed, the transmission could be easily scripted; hence no subject controllers were required for the airborne experiment. The pilots participated as single operators with no crew.

2. Subject Quantity, Total Participation Time, and Training Time

An unequal number of subjects participated in the two experiments, and the length of subject participation (including training) was different as well. The subject pools for the ground-based and airborne experiments were 10 controllers and 48 pilots, respectively. To collect data efficiently, each pool was divided into groups, and each group participated together in the same scenarios.

In the ground-based experiment, two groups of five controllers participated simultaneously in separate but identically configured four-sector simulation 'worlds.' In each world, the sectors were staffed with one controller each plus one additional controller as the area supervisor. The controllers were available for a single, common eight work-day period. Therefore, training, data collection, questionnaires, and debrief were scheduled during this period. Training was accomplished over three days, including 10 hours of hands-on training. This time period included additional data runs external to the common M and S test matrices described here. Results from these additional runs will be reported in separate publications.

In the airborne experiment, groups of 12 pilots participated in three-day sessions. Four sessions were scheduled over consecutive weeks, each with a different group of pilots, for a total of 48 pilots. Two pilots were retroactively excluded from data analysis because of lack of recent flying experience, resulting in a pool of 46 pilots for data analysis. The pilots' qualifications are described in Ref. 23. Since the pilots served as subject participants during their days off from flying the line, their maximum availability was three days onsite at Langley. Therefore, all training, data collection, questionnaires, and debriefing had to be completed during this time. The pilots received one day of training, including five hours of hands-on training, supplemented by pre-mailed reading material. During data runs, each group of pilots simultaneously flew their aircraft in shared scenarios, i.e., they piloted 12 of the several hundred aircraft operating in the experiment airspace. The remaining aircraft were automated, i.e., 'flown' by pilot model software, and they were not considered a valid source of experimental data for this HITL experiment.

3. Aircraft Monitored by Subject Participants

Even though the two experiments had identical aircraft populations, the subset of aircraft managed by the controllers and pilots were necessarily different from one another. This necessity resulted from different role assignments: controllers managing sectors of airspace and pilots flying individual aircraft. In the ground-based experiment, the controller/automation team managed separation for all several hundred aircraft when they were passing through the four ATC test sectors shown in the center of Fig.5 (ZKC90, ZKC98, ZID80, and ZID81). In the airborne experiment, each pilot flew one aircraft per scenario, and therefore only 12 of the hundreds of aircraft involved human pilots. They flew across the experimental airspace region, shown in Fig. 5, which was sized to enable 30 minutes of continuous flight. Since sector boundaries have no significance in the airborne concept, the ATC sectors played no role in the airborne experiment other than to provide a common reference for joint data analysis. Since the four sectors constituted only a subset of the total experimental airspace region, portions of the aircraft trajectories flown by the subject pilots lay outside of the four sectors controlled by the subject controllers.

4. Scenario Difficulty

The scenario difficulty experienced by the controller and pilot populations was defined by their dissimilar domains of control, i.e., within a sector for a controller and along a single trajectory for a pilot. Additionally, the

difficulty varied among controllers and among pilots. Each controller's experience in each run was defined by the unique flow patterns and local traffic densities of their sector, which were determined to be significantly non-uniform across sectors and across time. Therefore, the FAA-supplied controllers were rotated through the three more complex sectors (ZKC90, ZKC98, and ZID81) and the supervisor position. A recently retired controller worked the fourth sector (ZID80). Similarly, each subject pilot flew different aircraft in the same scenarios, which meant they had dissimilar experiences depending on their trajectory relative to the surrounding traffic (e.g., number and types of conflicts). Although they rotated through a common set of 24 aircraft across the complete set of 14 scenarios[*], in any one scenario (i.e., test condition), some pilots had busy flights whereas others had few or no conflicts. Both of these outcomes were a byproduct of the random scenario generation process and the naturally non-uniform characteristics of the traffic patterns in the selected airspace region.

Although the two experiments shared the same initial conditions of each scenario, it was not possible to guarantee that the two experiments would produce the same traffic conflicts. In fact, the probability was low that the same conflicts would appear, given that the operational concepts are founded on flexible use of the airspace. Once a trajectory is changed for any reason, the scenario unfolds differently. For instance, since the controllers participated primarily in an exception-handling role, they were less apt to adjust trajectories without a reason such as a conflict. However, the subject pilots in the airborne experiment were free to adjust their altitude and routing to their preference, provided that they created no conflicts in the process. In either concept, conflict resolutions early in the scenario altered encounters later in the scenario. In addition, the ground-based concept adhered to the cardinal altitudes, whereas the airborne concept permitted pilots to cruise at altitudes other than 1000-ft cardinal altitudes. The pilots of the airborne experiment frequently took advantage of this option, which had the unintended effect of increasing their exposure to conflicts since the aircraft without subject pilots remained at the cardinal altitudes.

V. Results and Discussion

Selecting metrics to compare two very different concepts tested in different simulation platforms was a significant challenge and is expected to see improvement with each subsequent conduct of common simulations. Some metrics support limited comparability whereas others, while important to present, should not be directly compared. The following categories of metrics are presented in this paper, and in the following subsections, each is introduced in terms of suitability for comparison.

A. <u>System Performance and Safety</u>: *Conflict detection; Time to LOS at initial detection and final clearing; Conflict resolution;* and *LOS events*
B. <u>Efficiency</u>: *Flight path deviation;* and *Schedule conformance*
C. <u>Subjective Assessments</u>: *Workload ratings;* and *Ratings of the operational concept and procedures.*

It is important to note that in this initial set of experiments, not all data from the simulations were considered valid for comparison. Comparisons between the two concepts were *restricted to the aircraft flown by the subject-pilots in the airborne experiment and, in the case of conflict metrics, to conflicts occurring in the test sectors staffed by air traffic controllers in the ground-based experiment.*

In the ground-based experiment, the performance of the confederate controllers and pseudo-pilots were excluded from analysis. Otherwise, since all aircraft were modeled at the same fidelity and were controlled by the same automation system, the metrics were collected for all aircraft across the airspace and the data for the aircraft that were used for the comparative analyses were filtered from the entire data set during post-processing During the data collection, the controllers had no knowledge that a subset of aircraft would be used for a special data analysis. Unlike the airborne experiment, all aircraft in the ground-based simulation were simulated in the same way by MACS, and the aircraft used for comparison in this paper received no added attention or treatment by the controllers, the automation, or the simulation systems. Aircraft were included in the analysis regardless as to whether a controller actively worked the aircraft or whether the aircraft was managed only by the ground-based automation without need for controller involvement.

In the airborne experiment, the only qualifying aircraft for data analysis were the 12 ASTOR aircraft in each scenario flown by the subject pilots. The remaining aircraft (batch ASTORs and TMX aircraft) provided the necessary background traffic but were not considered sufficiently capable and mature to provide data comparable to the piloted ASTORs. The pilot model software for the batch ASTORs was updated with many new capabilities needed to conduct the experiment, but its performance with these capabilities was not sufficiently robust for

---

[*] Two different sets of 12 aircraft were flown by pilots in the M and S scenarios, respectively.

analysis. The TMX aircraft used conflict resolution algorithms that were significantly different and less mature than those used in AOP. The inclusion of these aircraft in the simulation was necessary to reach the high traffic density throughout the large experimental airspace. During the data collection, the subject pilots were not aware of any differences between aircraft in the simulation, although they were aware that only 12 pilots participated at a time and that most of the aircraft were automatically controlled.

A. System Performance and Safety

System performance and safety results are presented in the areas of conflict detection, conflict resolution, and LOS events for the M scenarios. The conflicts in the S scenarios have not yet been analyzed in detail, but are expected to follow similar patterns as the 2.0x M scenarios. In general, metrics of conflict detection and conflict resolution can be loosely compared between the two experiments, while taking note that significant differences existed in trajectory modeling and automation design between the two simulations. The quantity of LOS events cannot be directly compared, nor can events be compared between specific aircraft pairs in the two experiments, because each LOS was unique to its environment and preceding events. However, areas of commonality can be explored in the general types of causal factors that led to these events.

1. Conflict Detection

The number of conflicts was counted within the common portion of the two experiments' reportable data, i.e., conflict detections involving the 12 aircraft flown by pilot participants in the airborne experiment and for which the predicted LOS occurred within the four sectors staffed by the controller participants in the ground-based experiment. The data presented throughout this section only include conflicts that persisted for at least 12 seconds, a traditional indication of conflict stability used in AOL data analysis. In addition, a repeated conflict detection involving the same pair of aircraft in which the gap between active reported times was less than 90 seconds was counted as single conflict.

As shown in Fig. 10, the distribution of conflicts between the four ATC sectors was similar for the two experiments. By far, the airspace region with the most
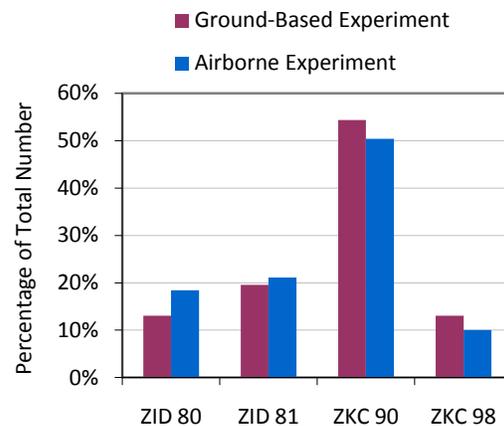
Figure 10. Distribution of conflicts within the ATC sectors. Conflict location is defined by the predicted LOS location.

conflicts was Kansas City Center sector 90. Table 3 shows the total number of conflict detections across all runs and their distribution by sector. Sector ZKC 90 had over 30 percent more conflicts than any other sectors. These conflicts were further analyzed, and it was determined that many of these resulted from the large number of descending and climbing aircraft in this sector.

The mean number of conflicts detected in each of the four M scenarios is shown for the ground-based and airborne experiments in Fig. 11, and the descriptive statistics are in Table 4.

There are at least three notable effects. (1) The airborne experiment recorded a greater number of

Table 3. Total conflicts detected in the four ATC sectors across all M scenario runs. Note that the ground-based experiment included half the number of runs of the airborne experiment.

Ground-Based Experiment							
	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total
M1	1.5x	No	4	0	5	0	9
M2	2.0x	No	4	9	27	2	42
M4	1.5x	Yes	3	0	5	1	9
M3	2.0x	Yes	1	9	13	9	32
		Overall	12	18	50	12	92
	Overall Percentage		13%	20%	54%	13%	100%
Airborne Experiment							
	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total
M1	1.5x	No	19	13	40	9	81
M2	2.0x	No	15	18	24	7	64
M4	1.5x	Yes	18	12	33	3	66
M3	2.0x	Yes	12	25	62	9	108
		Overall	64	68	159	28	319
	Overall Percentage		20%	21%	50%	9%	100%
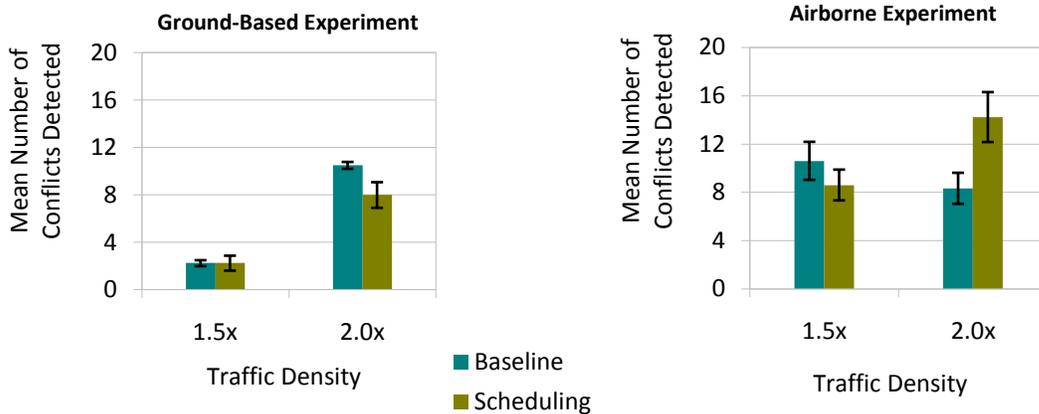
Figure 11. Mean number of conflicts detected in the M scenarios. Error bars are +/- 1 standard error of the mean.

detected conflicts than the ground-based experiment. (2) The ground-based experiment data shows an unusually strong effect of traffic density on conflicts detected. (3) The airborne experiment data for the baseline runs (M1 and M4) show an unexpected trend for traffic density.

Understanding the reasons for these effects is important, because automated conflict detection is at the center of future separation assurance concepts. A detailed data analysis is underway to gain insights into the reasons for the effects described above, but the complete results are not available at this time. Preliminary data and observations point to the following potential reasons for the behavior noted above.

(1) The airborne experiment recorded a greater number of detected conflicts than the ground-based experiment The ground-based experiment used a deterministic conflict probe that examined the predicted locations of each aircraft for potential violations of the preset separation minima. When the predicted flight state of one aircraft included a climb or a descent, a heuristic was applied that compared the current, the predicted, and the target altitudes for both aircraft to assess whether an additional buffer needed to be applied to the separation standard. If necessary, such a buffer was added for each transitioning aircraft. The airborne experiment used a different conflict detection algorithm that, in addition to the traditional separation buffers, added buffers around trajectory segments.[20] The purpose of applying these additional buffers was to account for variations in the predictability of aircraft position on these different segments, e.g. transition segments (climbs and descents) vs. level segments, and turn segments vs. constant track segments. The method worked by expanding each line segment of a trajectory in the lateral, vertical, and along-path dimensions into a volume that encompassed the total aircraft position uncertainty defined for that segment type. The conflict detection function then determined whether the volume edges of the ownship trajectory were predicted to lose separation with the volume edges of each traffic aircraft trajectory. It is believed that the trajectory segment buffers used in the airborne experiment likely resulted in a more conservative conflict detection behavior, i.e., a larger number of false alerts. A preliminary analysis indicated that approximately half of the conflicts in the airborne experiment (163 of 319) involved one or both aircraft on a transition segment with a predicted vertical separation greater than 1000 ft (indicating the vertical trajectory buffers caused the conflict to be detected). The most frequent occurrences were in the scenarios with the largest number of conflicts detected (M1 and M3). Excluding these conflicts, the mean numbers of conflicts detected were 2.25, 2.75, 3.38, and 2.88 for scenarios M1, M2, M4, and M3, respectively. As seen in Table 4, these counts are similar to those for the ground-

Table 4. Descriptive statistics for the conflicts detected in the M scenarios.

	Density	Schedule	Ground-Based Experiment			Airborne Experiment		
			Mean	SD	StdErr	Mean	SD	StdErr
M1	1.5x	No	2.25	0.50	0.25	10.60	4.46	1.58
M2	2.0x	No	10.50	0.58	0.29	8.33	3.64	1.29
M4	1.5x	Yes	2.25	1.26	0.63	8.60	3.60	1.27
M3	2.0x	Yes	8.00	2.16	1.08	14.23	5.85	2.07

American Institute of Aeronautics and Astronautics

based experiment at 1.5x traffic density.

(2) The ground-based experiment data shows an unusual strong effect of traffic density on conflicts detected

This effect correlates with the observation that the 2.0x scenarios resulted in much higher local complexities than the 1.5x scenarios. These complexities were related to additional traffic from certain flows rather than the overall density increase that resulted in more conflicts than the n-squared increase in conflicts would normally predict. Table 3 indicates that the increase in detected conflicts was not uniform across all sectors. The subjective data presented later in this paper indicate higher workload and more safety and acceptability concerns with the 2.0x scenarios than with the 1.5x scenarios specifically for the ZKC sectors. Data analysis is underway to categorize the conflicts further and analyze the complexity of the local areas to gain further insight into this issue.

(3) The airborne experiment data for the baseline runs show an unexpected trend for traffic density

The airborne experiment data indicate that more conflicts were detected during the 1.5x baseline runs (M1) than for the 2.0x baseline runs (M4). However, when those conflicts are excluded that involved one or both aircraft on a transition segment with a predicted vertical separation greater than 1000 ft, the trend changes towards the expected behavior. Therefore, the same explanation that applied to the first point above may apply here as well. However, it is still unclear why the 2.0x baseline scenarios did not result in more conflicts with transitioning aircraft than the 1.5x baseline scenarios. Therefore, a more detailed analysis is required to understand this issue entirely.

2. Time to LOS at Initial Detection and Final Clearing

The time to the predicted LOS when conflicts are first detected and finally cleared are two metrics of interest to system performance and safety. The time of first detection is determined primarily by the look-ahead parameter of the conflict detection function, which for both experiments was set to 10 minutes, but also by prediction uncertainties and aircraft maneuvers that may delay detection. In terms of system performance and safety, it is desired to detect conflicts early enough to have sufficient time to resolve them. The time of final clearing is governed by the time required to request, compute, load, and execute a resolution. It also is governed by any repeated conflicts between the same aircraft pair, as only the final clear time is registered. In the airborne concept, the time at which the conflict is displayed to the pilot is also a factor, because an aircraft with higher priority will receive the alert later (if still necessary), giving the burdened aircraft a chance to solve the conflict first. In terms of system performance and safety, it is desired that conflicts be resolved sufficiently far in advance of the predicted LOS to minimize the need for urgent tactical actions by the controller or pilot.

The mean times-to-LOS at first detection and final clearing are shown in Fig. 12 for the M scenarios, and Table 5 shows the descriptive statistics. The data included in these plots were derived from the same subset of the total conflicts included in the previous conflict count analysis. In both data sets, the mean times for initial detection and final clearing were greater than five minutes to the predicted LOS, indicating generally adequate mean system performance. A greater effect of traffic density is apparent in the ground-based experiment data than for the airborne experiment, as indicated by slight reductions in the mean time-to-LOS at first detection and final clearing, as traffic density was increased from 1.5x to 2.0x. However, overall, the conflicts in the ground-based experiment were detected and resolved earlier than those in the airborne experiment. Nearly all subject-aircraft conflicts in the airborne experiment occurred with TMX aircraft, and a review of data revealed software faults in TMX that, in
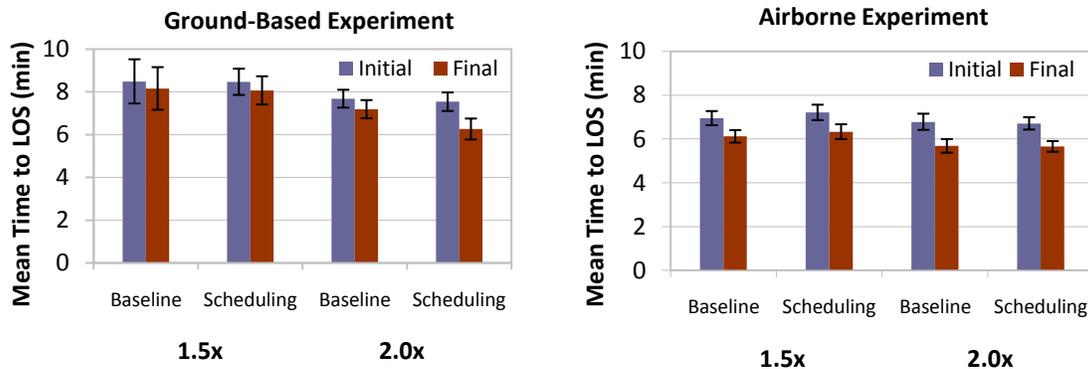

Figure 12. Mean initial and final times to LOS for conflicts in the M scenarios. Error bars are +/- 1 standard error of the mean.

**Table 5.   Descriptive statistics for the initial and final times to LOS for conflicts in the M scenarios.**

			Initial Detection				Final Clear			
							Ground-Based Experiment			
	Density	Schedule	N	Mean	SD	StdErr	N	Mean	SD	StdErr
M1	1.5x	No	9	8.49	3.10	1.03	9	8.16	2.99	1.00
M2	2.0x	No	42	7.68	2.71	0.42	42	7.19	2.75	0.42
M4	1.5x	Yes	9	8.47	1.84	0.61	9	8.07	1.98	0.66
M3	2.0x	Yes	32	7.54	2.46	0.43	32	6.26	2.78	0.49
							Airborne Experiment			
			Initial Detection				Final Clear			
	Density	Schedule	N	Mean	SD	StdErr	N	Mean	SD	StdErr
M1	1.5x	No	81	6.95	2.91	0.32	81	6.12	2.57	0.29
M2	2.0x	No	64	6.78	2.96	0.37	64	5.68	2.48	0.31
M4	1.5x	Yes	65	7.21	2.84	0.35	65	6.33	2.72	0.34
M3	2.0x	Yes	109	6.71	2.95	0.28	109	5.66	2.55	0.24

many cases, delayed the timely detection and resolution of conflicts by either aircraft in the conflict. These faults included the broadcast of incorrect target-state altitudes, which significantly affected the AOP's accuracy in conflict detection, and incorrect filtering out of conflicts by TMX (for reducing computational load to enable the large total aircraft count), which reduced TMX aircraft participation in detecting and resolving conflicts. These effects resulted in some conflicts being detected and resolved with less time than intended by the system design. As will be discussed in the next section, these software faults also resulted in some instances of LOS. Finally, as expected, the inclusion of schedule constraints had little effect on the timing of the detection and resolution of conflicts.

3. Conflict Resolution
   Figure 13 shows the distribution of conflict resolutions among the four ATC sectors. The accompanying data is in Table 6. As expected, the data parallel the distribution of the conflicts themselves (Fig. 10) and indicate that the airspace corresponding to sector 90 was the busiest airspace for trajectory changes in both experiments. It should be noted, however, that the aircraft may not have actually been in the sector of the predicted LOS location when the conflict resolution occurred, and the resolution maneuver itself may not have always been within the sector. Sector boundaries are not part of the airborne concept and were not, therefore, represented to the pilots or the AOP in the experiment.
   Comparing Table 3 (conflict detections) to Table 6 (conflict resolutions), there is an approximately equal number of conflicts detected and conflict resolutions in the ground-based experiment. Pending further analysis, small
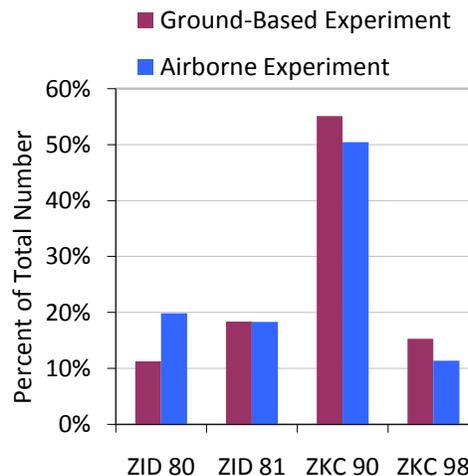

Figure 13. Distribution of resolutions for conflicts occurring within the ATC sectors. Conflict location is defined by the predicted LOS location.

Table 6. Total conflict resolutions for conflicts in the four ATC sectors across all M scenario runs. Note that the ground-based experiment included half the number of runs of the airborne experiment.

	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total
Ground-Based Experiment							
M1	1.5x	No	4	0	4	0	8
M2	2.0x	No	3	9	30	2	44
M4	1.5x	Yes	3	0	6	1	10
M3	2.0x	Yes	1	9	14	12	36
		Overall	11	18	54	15	98
		Overall Percentage	11%	18%	55%	15%	100%
Airborne Experiment							
	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total
M1	1.5x	No	11.8	10.4	33.6	8.2	64.0
M2	2.0x	No	10.0	9.6	17.6	4.2	41.4
M4	1.5x	Yes	7.4	5.2	27.0	2.0	41.6
M3	2.0x	Yes	11.6	12.4	25.4	9.0	58.4
		Overall	40.8	37.6	103.6	23.4	205.4
		Overall Percentage	20%	18%	50%	11%	100%

differences can likely be attributed to false alerts, conflicts that were resolved as a result of a trajectory change intended for schedule management or to solve a different conflict. In the airborne experiment, there were approximately 55 percent more conflicts detected than resolutions. Several factors account for this difference, including false-alert conflicts that cleared on their own, resolution maneuvers that resolved more than one conflict at a time, and conflicts resolved by the traffic aircraft.

Figure 14 presents the mean conflict resolution count for the ground-based and airborne experiments, using the same subset of data described for the conflict detection analysis above (i.e., M scenarios only, conflicts involving subject-piloted aircraft, with predicted LOS in ATC sectors, having at least 12 second duration, and separated by at least 90 seconds from other conflicts same-run/same-pair aircraft). Table 7 shows the resolutions by ATC sector and type of resolution in the ground-based experiment, and Table 8 shows similar data for the airborne experiment. The plot indicates the total number of resolutions for each experimental condition, as well as the proportion of strategic and tactical resolutions. The total number of resolutions mirrors the conflict detection trends depicted in Fig. 11, as expected, although fewer resolutions were recorded in the airborne experiment relative to the ground-based experiment during the 2.0x traffic density conditions. This result is consistent with the distributed nature of the airborne concept, in which the traffic aircraft resolves some of the conflicts. In the airborne experiment, the data reflect only the resolutions of the subject-piloted aircraft and not those of the traffic aircraft, whereas the ground-based experiment data reflect resolutions by both aircraft.
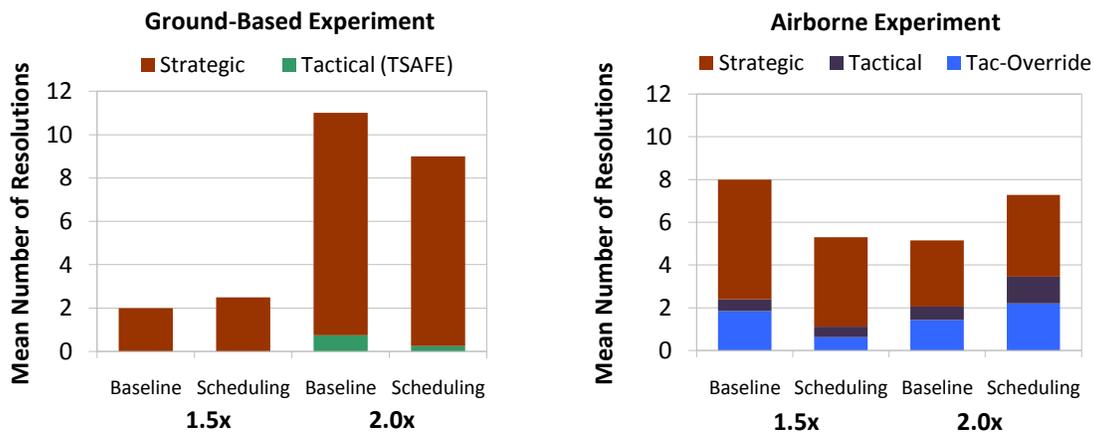
Figure 14. Breakdown of conflict resolution type for conflicts in the M scenarios. The data only include conflicts involving subject-piloted aircraft with LOS predicted to occur within the ATC sectors.

Table 7. Resolutions for conflicts in the four ATC sectors in the ground-based experiment.

	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total	Mean
Tactical Resolutions (TSAFE)								
M1	1.5x	No	0	0	0	0	0	0
M2	2.0x	No	0	0	3	0	3	0.75
M4	1.5x	Yes	0	0	0	0	0	0
M3	2.0x	Yes	0	0	0	1	1	0.25
		Overall	0	0	3	1	4	
Strategic Resolutions								
	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total	Mean
M1	1.5x	No	4	0	4	0	8	2.00
M2	2.0x	No	3	9	27	2	41	10.25
M4	1.5x	Yes	3	0	6	1	10	2.50
M3	2.0x	Yes	1	9	14	11	35	8.75
		Overall	11	18	51	14	94	

Table 8. Resolutions for conflicts in the four ATC sectors in the airborne experiment.

	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total	Mean
Tactical Override Resolutions								
M1	1.5x	No	3	3	8	0	14	1.85
M2	2.0x	No	5	2	3	1	11	1.42
M4	1.5x	Yes	2	2	0	1	5	0.62
M3	2.0x	Yes	2	3	7	5	17	2.20
		Overall	12	10	18	7	47	
Tactical (Non-Override) Resolutions								
	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total	Mean
M1	1.5x	No	0	3	0	1	4	0.55
M2	2.0x	No	2	2	1	0	5	0.62
M4	1.5x	Yes	2	2	1	0	5	.050
M3	2.0x	Yes	1	4	2	3	10	1.25
		Overall	5	11	4	4	23	
Strategic Resolutions								
	Density	Schedule	ZID 80	ZID 81	ZKC 90	ZKC 98	Total	Mean
M1	1.5x	No	8	4	24	7	43	5.60
M2	2.0x	No	3	5	13	3	24	3.10
M4	1.5x	Yes	5	1	25	1	32	4.18
M3	2.0x	Yes	8	5	15	1	29	3.82
		Overall	24	15	77	12	128	

In the ground-based experiment, tactical resolutions were not observed in the 1.5x traffic density condition and were a small proportion of resolutions in the 2.0x traffic density condition. In the airborne experiment, tactical resolutions occurred in each test condition and were a larger proportion overall. The airborne system had two tactical modes: tactical and tactical override. The tactical (non-override) mode was a normal mode used for conflict resolution whenever the auto-flight system was not fully coupled to the FMS or was predicted to decouple within AOP's conflict resolution look-ahead horizon of 20 minutes. As such, tactical (non-override) resolutions were not necessarily an indication of degraded system performance or safety. The tactical override mode, as its name implies, would override any strategic resolutions on the pilot's display, and the pilot was compelled by procedures to execute either the lateral or vertical tactical resolution. This mode is more comparable to the ground-based concept's tactical mode based on the TSAFE algorithm. However, the airborne tactical override mode was triggered when time-to-LOS was less than five or four minutes, depending on the aircraft's right-of-way, whereas TSAFE was triggered when time-to-LOS was less than three minutes. The greater proportion of tactical override resolutions in the airborne experiment data is consistent with the earlier trigger time.

4. Loss of Separation Events

Losses of separation occurred in both experiments.[*] However, of the comparable dataset used in the preceding conflict analysis (i.e., the aircraft piloted in the airborne experiment with LOS predicted to occur - or in this case actually occurring - within the ATC sectors of the ground-based experiment), no LOS events occurred in the ground-based experiment runs for these aircraft in these sectors, and six LOS events occurred in the airborne experiment runs. Simulation faults contributed to all but one of these LOS events and will be discussed below.

As this paper focuses primarily on comparable results, other LOS events will not be discussed here in detail. In the ground-based experiment, other LOS events in the ATC sectors involved aircraft not piloted in the airborne experiment. These will be discussed below in general terms to address some of the contributing factors, and a detailed analysis will be available in future publications.[23] In the airborne experiment, subject aircraft were involved in four additional LOS events occurring outside the ATC sector boundaries. These will be included in the discussion below because the contributing factors were similar to those within the ATC sectors.[†] Not discussed for either experiment are LOS events occurring outside the ATC sectors in the ground-based experiment or involving only non-piloted aircraft in the airborne experiment, as they did not have the full benefit of the respective human-automation concepts.

LOS in HITL simulations pinpoint problems that need to be addressed and are rich and important sources of information. The value of a LOS analysis is to assess the causal factors so that improvements in technology and/or procedures can be inserted back into the concept. The following discussions for each experiment illuminate the wide variety of factors that resulted in LOS, from complex traffic flow interactions to simulation software faults to human procedural errors.

Ground-Based Experiment LOS

Of the subset of aircraft used for comparison throughout this paper, i.e., the piloted ASTORs in the airborne concept simulation, none were involved in any losses of separation in the ground-based experiment. However, other elements of the scenarios included extremely complex traffic flow interactions, often involving climbing and descending aircraft and provoking hard-to-resolve short-term conflicts. These situations stressed the function allocation concept and sometimes resulted in LOS events in some sectors involving aircraft not considered for data comparison in this paper. The results are included in other publications[23] that discuss additional findings related to the other aspects of the ground-based study.

The primary reason for a LOS in the ground-based approach is usually very late conflict detection related to complex interactions of various factors. For example, in the M scenarios at 2.0x traffic density, dense departure streams from the St. Louis airport (centrally located in the rectangular experiment airspace) climbed into dense arrival and overflight traffic in the northern part of ZKC90 and the southern part of ZKC98. The aircraft were on steep climb trajectories and were handed off from a confederate controller's airspace into the test airspace. Since aircraft control systems do not try to maintain a predefined vertical flight profile on climb trajectories, the prediction accuracy for climb trajectories is generally much worse than for en route or arrival aircraft. Therefore, conflicts were often detected very late, with less than four minutes to the predicted LOS. This short horizon was too close for the trajectory-based automation to issue a trajectory amendment automatically. The controllers frequently wanted to stop the climbing aircraft until clear of the conflicting traffic and then resume the climb. To accomplish this, they sometimes uplinked a new trajectory with a new altitude, and other times they issued voice commands. Due to the situation's urgency, the controllers usually did not use the automation tools to trial-plan and conflict-probe the altitude assignment issued by voice. They were also sometimes unable to easily find an altitude that cleared the 7 nmi buffer used in the trial-plan conflict detection tool. While this was happening, TSAFE would at times also issue a heading to one or both of the aircraft, requiring the controller to then deal with the off-track situation and re-evaluate the altitude he or she had issued, which may not have originally been conflict-free. Dealing with such situations increases controller workload significantly, and draws the controller's attention to a specific problem. Even more challenging was the issue that, due to the density of the departure flow, two or three of these conflicts were sometimes flagged simultaneously to the controller, with little time to solve the problems. In this study, the controllers were not able to stop the automation from issuing a TSAFE heading, and an even more complex situation could arise, sometimes leading to a LOS.

In the future, various changes will be investigated to address these kinds of problems. For example, in addition to heading changes, TSAFE will be allowed to send altitude changes to the aircraft. Trial-plan altitudes will be pre-probed so that controllers will not have to find a clear altitude by trial and error. Controllers will have a means to

---

[*] The criteria for loss of separation were a closest point of approach within 5 nmi and 800 ft.
[†] ATC sectors were not modeled in the airborne experiment as they were not relevant to the concept.

stop TSAFE from sending any changes. Traffic flow managers will be available to precondition the departure flows from the confederate controllers' airspace, if necessary, to manage the complexity inside the sectors.

Airborne Experiment LOS

In the airborne experiment, six losses of separation involving subject-piloted aircraft occurred in the ATC sector airspace of the ground-based experiment, and four occurred elsewhere. Of the six, two were "proximity events" with closest points of approach greater than 4.5 nmi, and four were "operational errors" with closest points of approach less than 4.5 nmi. Of the remaining four, three were operational errors. Two LOS events (one inside and one outside the ATC sectors) occurred to the same subject-piloted aircraft within a brief span of time. Every LOS event was between a subject-piloted aircraft and a TMX aircraft. Short conflict detection time was a factor in all 10 LOS events and was caused by a variety of factors, including simulation faults and pilot error, as described below.

An analysis of each of the 10 LOS events was performed to determine the primary and contributing factors leading to the event. Figure 15 shows the primary factors. Five were attributed to automation or simulation faults, four to a combination of automation/simulation faults and pilot error, and one to pilot error alone. In most of the cases attributed to automation, where the automation failed to provide the pilot with adequate guidance to prevent the LOS, the root cause was determined to be the TMX software faults discussed earlier in the time-to-LOS analysis and in more detail here.

In six of the nine events attributed partially or fully to automation (blue and red in Fig. 15), incorrect filtering out of conflicts by TMX caused the TMX aircraft to never see the ASTOR aircraft prior to the LOS. The filtering was included to reduce computational load in TMX to enable the large aircraft count required by the scenarios. As a result of the incorrect filtering, TMX aircraft could not take its own independent action to resolve the conflicts as intended in the concept. Four of the five LOS events attributed solely to automation (blue in Fig. 15) were the result of an additional software error in TMX in which wrong target altitudes were broadcast. This error caused the AOP to predict that no conflict existed, a false prediction that persisted until the actual position of the TMX aircraft was close enough that the incorrect projected path was inside the LOS criteria and the conflict was detected. As a result, the pilots were alerted with only zero to 20 seconds notice, and because no conflict-free

Figure 15. Primary factors leading to the subject-pilot LOS events in the airborne experiment. All 10 events from the experiment are shown. Six of the events occurred within the airspace of the four ATC sectors.

tactical solutions were available, not enough time or guidance was provided to prevent the actual losses of separation. The fifth LOS event attributed to automation was a result of the TMX aircraft sending out incorrect TCPs; the pilot was alerted with 20 seconds warning, and no tactical override solution was available.

The four LOS events attributed to both automation and pilot error (red in Fig. 15) occurred in part because the automation was given faulty information, as described above, but also because the pilot took some procedurally incorrect action to help cause the LOS. Two of the four cases were related to inaccuracy of the AOP's climb prediction for the traffic aircraft; the TMX aircraft's true climb rate was much less than the prediction, resulting in first detections of 19 and 20 seconds time-to-LOS. The automation did provide tactical override guidance, but the pilots delayed their execution actions, resulting in the two LOS events. In the other two LOS events, the pilots placed the auto-flight system in modes unsupported by the AOP, despite training and annunciation of the unsupported mode, resulting in first detections of 19 and 120 seconds time-to-LOS and eventual LOS.

The single LOS event attributed to pilot error alone was due to the pilot becoming task saturated and losing situation awareness. Against procedures, he descended into the LOS. All 10 LOS events are undergoing further review to determine the appropriate changes to automation capabilities, pilot procedures, and training to prevent recurrences.

Several design and procedural changes will be made based on the analysis of these LOS events. Simulation faults will be corrected to ensure that all aircraft are correctly receiving surveillance information and broadcasting the correct trajectory information, which should significantly reduce the number of late conflict alerts to the pilot. AOP support for flight modes will be increased to reduce or eliminate the number of unsupported conditions, and pilot procedures for handling unsupported modes will be improved. Tactical guidance for short term conflicts will be improved to eliminate any automation delays in providing resolutions and to ensure a solution is always available. Vertical guidance cues will be redesigned to be more obvious and easier for the pilot to follow. Finally, procedures will change to make sure the pilot will not be distracted by the FMS when tactical conflict resolution maneuvers are required.
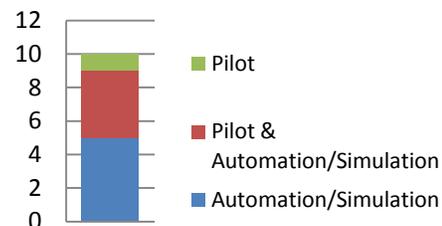
B. Efficiency

Efficiency results are presented in the areas of flight path deviation from the initial trajectory and schedule conformance. These metrics are comparable between the two experiments, although procedural differences between the pilots in one experiment and the controllers in the other have an impact on the results. Statistical analysis of each experiment's respective dataset was performed using an Analysis of Variance (ANOVA). Rank transformation was applied to the data when necessary to reduce the influence of outliers and stabilize variance across scenarios. Analyses for the ground-based experiment treated each individual aircraft as an independent data point, whereas for the airborne experiment, data were averaged over the 12 pilots within each group (recall that four groups of 12 pilots participated in the airborne experiment).

1. Flight Path Deviation (overall)

The M scenarios facilitated an investigation of two independent variables: traffic density (1.5x and 2.0x), and the presence of scheduling (RTA/STA assignment, no RTA/STA assignment). The S scenarios facilitated an investigation of the timing of trajectory changes due to scheduling constraints: a single unchanged RTA/STA assignment, a second revised RTA/STA assignment for all scheduled aircraft occurring at dispersed times, and a second revised RTA/STA assignment for all scheduled aircraft occurring at a synchronized time. For each experiment run, the lateral flight path deviation from the initial nominal trajectory, i.e. excess horizontal path, was computed for each subject-piloted aircraft. Table 9 presents descriptive statistics associated with these data for the entire sets of M and S scenarios. A negative value indicates that the trajectory actually flown during the run was shorter than the original trajectory for that aircraft. In the airborne experiment data, the large maximum deviation associated with the M scenarios (168.7 nmi) was the result of two aircraft that turned 360 degrees due to their circumstances. Excluding these two data points, the maximum flight path deviation was 37.0 nmi.

Table 9. Descriptive statistics for lateral flight path deviation (nmi) for the M and S scenarios.

Ground-Based Experiment						
Matrix	N	Mean	SD	Min	Med	Max
M	192	3.8	8.3	-2.4	0	44.7
S	432	4.9	10.1	-3.2	0	75.5
Airborne Experiment						
Matrix	N	Mean	SD	Min	Med	Max
M	366	2.3	11.1	-7.9	0	168.7
S	270	2.1	4.2	-0.1	0	23.8

Effect of Traffic Density and Schedule Assignment (M Scenarios)

Descriptive statistics associated with the four M scenarios are shown in Table 10.[*]

Table 10. Descriptive statistics for lateral flight path deviation (nmi) in each M scenario.

	Traffic Density	Schedule Assignment	N	Mean	SD	Min	Med	Max
Ground-Based Experiment								
M1	1.5x	No	48	2.2	7.8	-1.2	0	38.9
M2	2.0x	No	48	5.0	8.3	0	2.7	40.2
M4	1.5x	Yes	48	1.9	7.1	-2.4	0	44.7
M3	2.0x	Yes	48	5.9	9.5	0	0.4	35.6
Airborne Experiment								
M1	1.5x	No	91	0.8	3.7	-6.9	-0.1	13.3
M2	2.0x	No	92	3.3	12.1	-0.9	0	104.1
M4	1.5x	Yes	92	0.8	3.0	-7.9	-0.1	12.5
M3	2.0x	Yes	91	4.3	17.8	-1.1	0.4	168.7

In the ground-based experiment data, a significant main effect was observed in the traffic density manipulation ($p = 0.004$), resulting in larger lateral flight path deviations in the 2.0x traffic density (see upper two histograms in Fig. 16). No statistically significant difference was found in the scheduling manipulation ($p = 0.791$), and the interaction effect was also not significant ($p = 0.617$).

---

[*] In the airborne experiment, two data points were excluded from the flight path deviation analysis. In M1, an aircraft prematurely exited the experiment airspace, and in M3, an ASTOR software crash prematurely terminated another aircraft's flight.

American Institute of Aeronautics and Astronautics
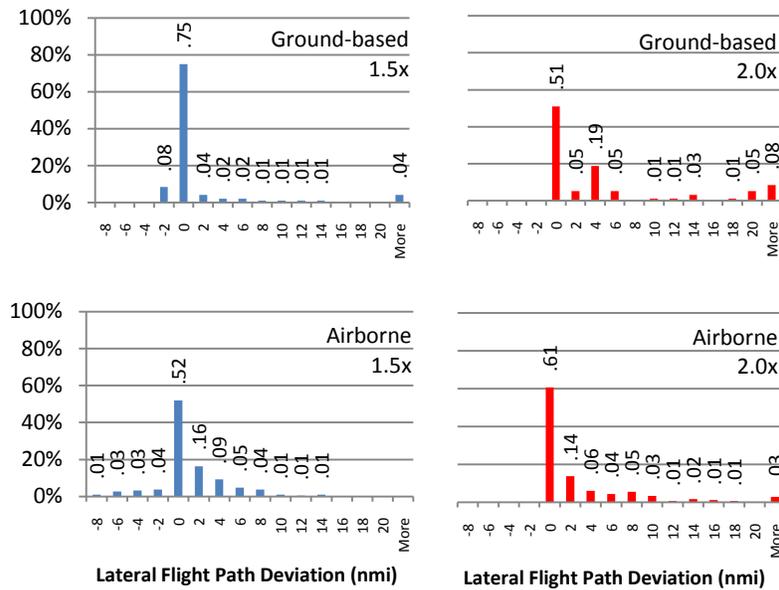
Figure 16. Histograms of lateral flight path deviation for traffic densities of 1.5x and 2.0x.

In the airborne experiment, similar to the ground-based experiment results, there was no statistically significant difference in the mean lateral flight path deviation in the scheduling manipulation ($p = 0.647$). However, the traffic density effect was found to be significant ($p < 0.001$), indicating that increasing the traffic density from 1.5x to 2.0x increased the mean lateral flight path deviation, which can also be seen in the distributions of the two lower histograms in Fig. 16. The interaction effect between scheduling assignment and traffic density was also not found to be statistically significant ($p = 0.576$).

Effect of Timing of the Trajectory Change Event (S scenarios)

Descriptive statistics associated with the three S scenarios are shown in Table 11.

For the ground-based experiment, no statistically significant differences were observed between the three scheduling conditions. This is likely due to the fact that, as discussed further in the schedule conformance analysis, the controllers did not have enough time during the 15-minute scenarios to always implement the trajectory change needed to absorb the delay from the second STA assignment. The ground-based experiment's distribution of the mean lateral path deviation for the different rescheduling conditions is shown in comparison to the airborne experiment's distributions in Fig. 17. The distribution of the mean lateral path deviation from the ground-based experiment indicates that the three scenarios were worked by the controllers in a similar fashion.

For the airborne experiment, in contrast to the data from the ground-based experiment, the trajectory change event was found to have a statistically significant effect on the mean lateral flight path deviation ($p = 0.005$). The result was expected, because the event was designed to require a path stretch to absorb an arrival delay. Tukey simultaneous pairwise comparisons revealed that the mean lateral flight path deviation was significantly lower for scenario S1 compared to scenarios S2 ($p = 0.038$) and S3 ($p = 0.005$). However, there was no statistically significant difference between scenarios S2 and S3 ($p = 0.589$). Therefore, the inclusion of a second RTA, triggering another trajectory change intended to absorb the new delay, increased the lateral flight path deviation as expected, but the timing of this event did not have a significant effect.

Table 11. Descriptive statistics for lateral flight path deviation (nmi) in each S scenario.

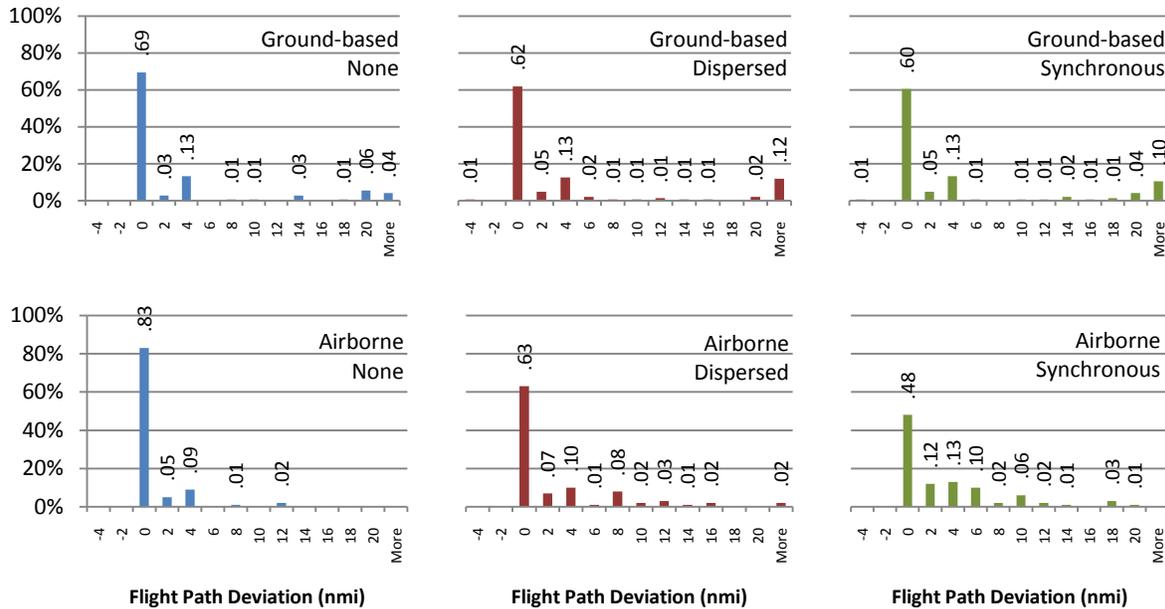|	Rescheduling	N	Mean	SD	Min	Med	Max
Ground-Based Experiment							
S1	None	144	3.5	7.5	0	0	40.0
S2	Dispersed	144	5.3	10.5	-3.2	0	40.8
S3	Synchronous	144	5.9	11.8	-3.2	0	75.5
Airborne Experiment							
S1	None	92	0.7	2.0	-0.1	0	11.8
S2	Dispersed	89	2.7	4.9	-0.1	0	23.8
S3	Synchronous	89	3.1	4.6	-0.1	0.7	19.6

Figure 17.   Histograms of lateral flight path deviation in each condition of the S matrix.

2. Schedule Conformance (overall)

For each experiment run, the difference between an aircraft's last reported ETA at the metering fix and its last issued STA (ground-based terminology) or RTA (airborne terminology) at that fix was computed to represent an aircraft's end-of-run schedule conformance. Ideally an aircraft's value for this metric would be as close as possible to zero, indicative of a predicted on-time arrival at the metering fix: neither early (negative numbers) nor late (positive numbers). Overall descriptive statistics for these data are given in Table 12 for the M and S scenarios where there were schedule assignments. The median schedule conformance indicated high schedule conformance for both experiments in both M and S scenarios. In the S scenarios of the ground-based experiment, the observed large negative minimum value (and therefore the large standard deviation) was indicative of some aircraft not yet having been issued a delay maneuver by the controller prior to the termination of the short-duration scenario. In most or all of these cases, it was likely that the trajectory amendments would have been issued at a later time, as controllers often prioritize these actions by the aircraft scheduled to arrive the earliest. In the airborne experiment, the large negative minimum value also indicated that the delay maneuvers were not always executed before the run expired. Both indications were confirmed by visual review of selected controller and pilot playback files. Also of note, data for the ground-side analysis used fewer data points than shown in the Tables 10 and 11 of the flight path deviation analysis. This difference is because not all flows into the seven local airports were managed by one of the four controller positions. Two of the seven local airports with the lightest flows were not metered, and thus aircraft flying to those destinations did not receive STA messages.

Table 12. Descriptive statistics for M and S scenario schedule conformance (sec).

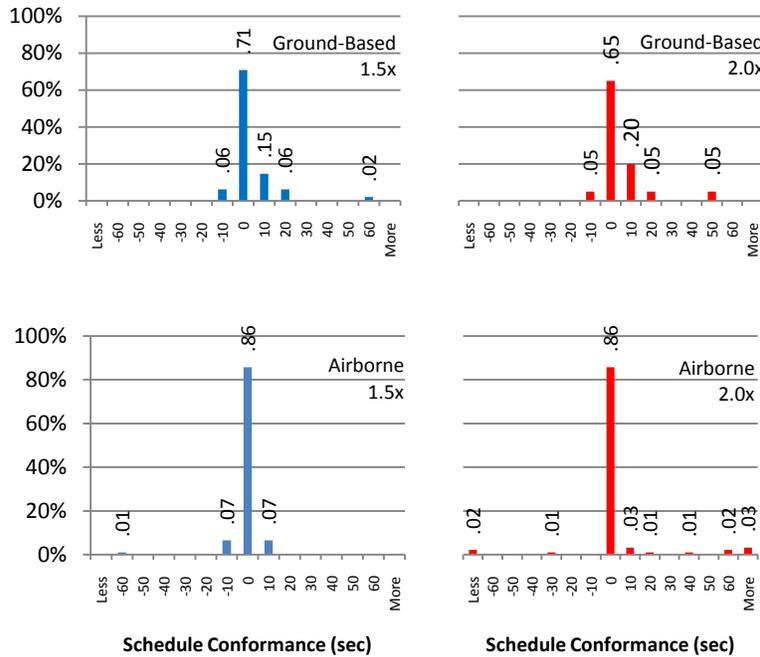| Ground-Based Experiment					
Matrix	N	Mean	SD	Min	Med	Max
M	88	5.0	10.8	-8	3	59
S	360	-73.1	133.6	-408	0	170
Airborne Experiment					
Matrix	N	Mean	SD	Min	Med	Max
M	183	8.3	87.6	-87	2	1170
S	270	-14.8	49.6	-322	-1	7

Effect of Traffic Density (M Scenarios)

Descriptive statistics associated with the M scenarios, in which traffic density was varied, are shown in Table 13.[*] Similar median schedule conformance, 3 seconds or less, is evident across traffic densities (1.5x and 2.0x) and across concepts (ground-based and airborne).

In the airborne experiment, the large mean and standard deviation at the 2.0x traffic density level were the result of the large maximum value (1170 sec), a single outlier in which an aircraft turned 360 degrees due to conflict circumstances and therefore fell significantly behind schedule. Excluding this data point, the maximum value was 94 seconds. From the histograms in Fig. 18, the higher peaks in the zero second bin (less than 5 seconds early or late) for the airborne experiment data are consistent with the airborne use of the FMS RTA functionality in which the auto-throttle actively controls speed to zero out RTA deviation. In the ground-based experiment, speeds were assigned to aircraft by the controller using the ground automation to ensure better matching of ground-based automation predictions to aircraft performance. Airborne RTA functionality was not used and the ground automation is designed to deliver aircraft within 15 seconds of their STA, which is considered sufficient for flow management purposes in the ground-based concept.

Table 13. Descriptive statistics for schedule conformance (sec) for the M scenarios.

Ground-Based Experiment						
Traffic Density	N	Mean	SD	Min	Med	Max
1.5x	48	4.2	10.0	-8	3	59
2.0x	40	6.0	11.8	-8	2.5	51
Airborne Experiment						
Traffic Density	N	Mean	SD	Min	Med	Max
1.5x	92	0.5	7.9	-65	2	11
2.0x	91	16.2	123.9	-87	2	1170

Figure 18.   Predicted schedule conformance for the 1.5x and 2.0x M scenarios.

Overall, across both traffic density conditions, the 95th percentile in the ground-based experiment was 21 seconds, i.e., 84 of 88 aircraft had a schedule deviation of 21 seconds or less, and the 99th percentile was 51, i.e., 87 of 88 aircraft. In the airborne experiment, the 95th percentile was 9 seconds, i.e., 173 of 183 aircraft had a schedule

---

[*] In the airborne experiment, a data point was excluded from the flight path deviation analysis. In M3, an ASTOR software crash prematurely terminated the aircraft's flight.

deviation of 9 seconds or less, and the 99th percentile was 74 seconds, i.e., 181 of 183 aircraft.   For both concepts, other than select outliers, the schedule conformance was quite good.
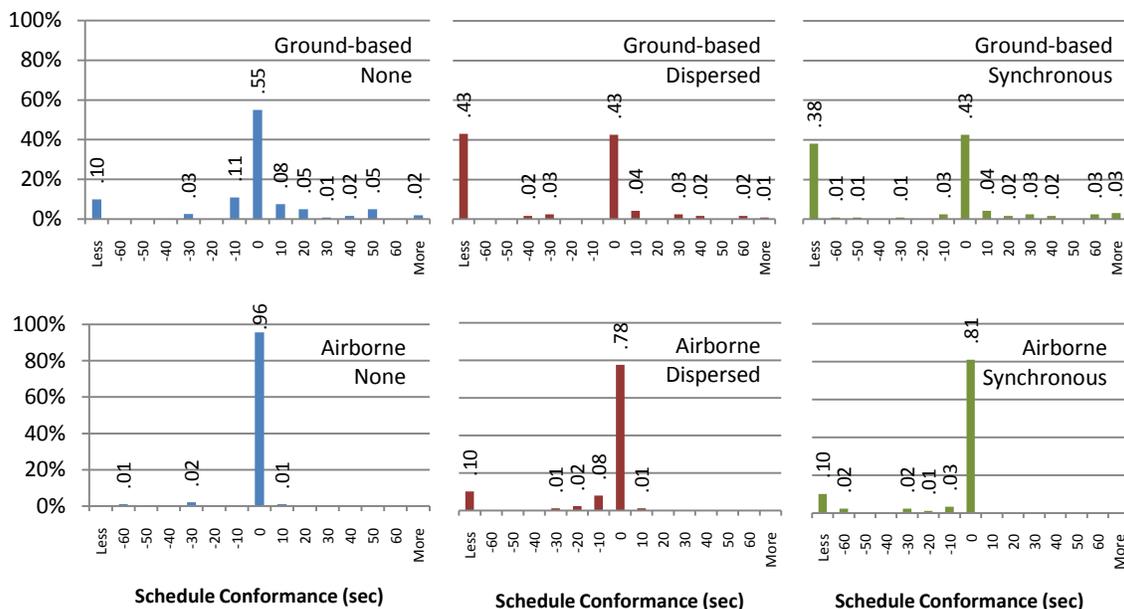
Analysis of the ground-based experiment data failed to obtain a significant effect of the traffic density manipulation ($p = 0.426$). In the airborne experiment data, traffic density was not found to have a statistically significant effect on predicted arrival delay ($p = 0.301$).

Effect of Timing of the Trajectory Change Event (S Scenarios)

Effects of the inclusion and timing of an arrival schedule revision necessitating delay maneuvers were investigated through the comparison of aircraft schedule conformance between scenarios S1 (initial STA/RTA only), S2 (revised STA/RTA at dispersed times), and S3 (revised STA/RTA at synchronous times). Descriptive statistics are presented in Table 14, and histograms are presented in Fig. 19. In the data of both experiments, a comparison of mean schedule conformance indicates clear differences between the non-rescheduled condition (none) and the rescheduled conditions (dispersed and synchronous). The airborne data had lower means and standard deviations as a result of less extreme outliers.

Table 14. Descriptive statistics for schedule conformance (sec) for the S scenarios.

Ground-Based Experiment						
Rescheduling	N	Mean	SD	Min	Med	Max
None	120	-5.5	44.7	-152	0	170
Dispersed	120	-117.7	151.9	-407	-1	70
Synchronous	120	-96.1	147.0	-407	-1	167
Airborne Experiment						
Rescheduling	N	Mean	SD	Min	Med	Max
None	92	-0.7	8.4	-63	1	6
Dispersed	89	-21.4	59.3	-286	-3	7
Synchronous	89	-22.7	60.2	-322	-3	3

Figure 19.   Predicted schedule conformance for the S scenarios.   Negative values represent early arrival predictions.

In the ground-based experiment's data analysis, a significant main effect of the schedule revision event on schedule compliance was also obtained ($p < 0.001$). Tukey simultaneous pairwise comparisons likewise showed greater schedule compliance in scenario S1 (no revision) compared to scenarios S2 ($p < 0.001$) and S3 ($p < 0.005$) with no difference found between S2 and S3 ($p = 0.625$). The fact that S1 was found to differ significantly from S2 and S3 is most apparently explainable by differences in the number of aircraft that were predicted to be more than two minutes early: five aircraft in S1, 48 aircraft in S2, and 41 aircraft in S3, respectively. An initial review of controller screen recordings indicate that this outcome was caused by controllers running out of time before the end of the S scenario. In other words, they did not have enough time to implement a second trajectory change needed to absorb the additional 4-5 minutes of delay in scenarios S2 and S3. Only two of these aircraft, however, were closer

than 25 minutes to their meter fix crossing point. Since controllers generally prioritized closer-in aircraft over those further out, they would have likely addressed these scheduling needs in time.

In the airborne experiment's data analysis, the trajectory change event was found to have a statistically significant effect on the mean predicted arrival time deviation ($p < 0.001$). Tukey simultaneous pairwise comparisons revealed that the mean predicted arrival delay was significantly lower for scenario S1 compared to scenarios S2 ($p < 0.001$) and S3 ($p < 0.001$). However, no statistically significant difference was indicated between scenarios S2 and S3 ($p = 0.968$). Therefore, the inclusion of a second RTA (indicating that a delay maneuver was required) had a significant effect on the mean predicted arrival delay, but the timing of this event did not.

C. Subjective Assessments

Results presented here reflecting subjective assessments of the participant controllers and pilots include workload and general concept acceptability. The subjective assessment results between the two experiments are not fully comparable, as they represent two different points of view of the concepts and are measured by instruments of different scales (to be described below). However, the results from the two concepts are presented together for limited comparison to provide a high-level context for assessing in general terms the human role in working with automation to accomplish a set of tasks resulting in safe passage of aircraft through the airspace. It should be noted that neither experiment collected subjective assessment data on the opposite role (pilots in the ground-based experiment and controllers in the airborne experiment), and that the controller and pilot participants were not briefed on the alternative concept of operations.

1. Workload Ratings (overall)

In the ground-based concept experiment, controllers provided a workload assessment after every run by completing portions of the NASA Task Load Index (TLX).[24] The original TLX has six scales that query participants about different aspects of their workload such as the mental demand of the task, their effort, and frustration. For all runs, the controllers completed the mental demand and time pressure scales. They rated these aspects of their workload on a seven-point scale, from "very low" to "very high." The second portion of the TLX that can be used to normalize participants' answers was not completed for this study.

In the airborne concept experiment, pilots provided a workload assessment after each scenario using the Modified Cooper-Harper (MCH) Subjective Workload Rating Scale.[25] Use of the MCH scale yields an overall workload rating ranging from "1" (indicating that the instructed task was very easy/highly desirable; operator mental effort was minimal; and desired performance was easily attainable) to "10" (indicating that the instructed task was impossible and could not be accomplished reliably). A rating of "3" indicates that the instructed task had a fair or mild difficulty level and that an acceptable level of operator mental effort was required to attain adequate system performance; therefore, an a priori decision was made that ratings of "3" or less would serve as an indication of an acceptable level of pilot workload.

Descriptive statistics from the controllers' TLX workload ratings in the ground-based experiment are shown in Table 15.[*] The controllers' mean workload ratings associated with the seven-point TLX scales indicate that they found the tasks performed during the ground-based concept's M scenarios to have a "below average" level of mental demand and a "low" degree of time pressure and found the tasks performed during the S scenarios to have a "medium" or "average" level of mental demand and a "below average" degree of time pressure. Descriptive statistics from the pilots' MCH workload ratings in the airborne experiment are shown in Table 16.[†] Similar to the ground-based experiment results, the pilots' mean workload ratings associated with the 10-point MCH scale indicate that they found the tasks performed during the airborne concept's M and S scenarios

Table 15. Descriptive statistics for controllers' TLX workload ratings in the ground-based experiment.

Mental Demand			
Matrix	N	Mean	SD
M	72	3.04	1.41
S	144	3.69	2.05
Time Pressure			
Matrix	N	Mean	SD
M	72	2.09	1.13
S	144	2.75	1.72

---

[*] Sample sizes of 64 and 128, respectively, were anticipated since each of the eight controllers was asked to provide workload ratings after completing each scenario (eight M scenarios and 16 S scenarios). Due to simulation mishaps, some runs were repeated giving additional data.

[†] Sample sizes of 368 and 276, respectively, were anticipated since each of the 46 pilots was asked to provide a MCH workload rating after completing each scenario (eight M scenarios and six S scenarios). Four pilots did not complete the MCH scale as requested, however, resulting in an absence of data.
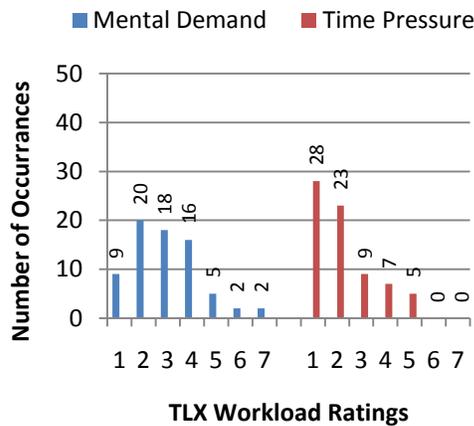
to have a difficulty level of "fair" or "mild" and that they felt an acceptable level of mental effort was required to attain adequate performance.

Distributions of the controllers' TLX mental demand and time pressure workload ratings associated with the M and S scenarios are shown in Figs. 20 and 21. For the ground-based concept's M scenarios, 38 percent of the controllers' TLX mental demand scale ratings had values of "1" or "2," and 67 percent of the time pressure scale ratings had values of "1" or "2." For the S scenarios, 41 percent of the controllers' mental demand ratings had values of "1" or "2," and 54 percent of the time pressure ratings had values of "1" or "2." Overall, these descriptive parameters indicate that the workload experienced by the controllers was within acceptable limits. The six mental demand ratings of "6" or higher recorded after the M scenarios indicate high load but were associated with runs that simulated 2.0x today's traffic, and individual differences are very evident. No time pressure ratings were given at this high end of the scale for the M scenarios.

Table 16. Descriptive statistics for pilots' MCH workload ratings in the airborne experiment.

Matrix	N	Mean	SD
M	340	2.41	1.80
S	253	2.14	1.33

During the ground-based concept's S scenarios, there were times that the controllers had many events co-occurring. The 18 mental demand ratings of "7," indicating very high load, were chosen through all three of the S scenarios (S1, S2, and S3) but always by controllers working the ZKC sectors. Similarly, the nine time pressure ratings of "6 " or above, indicating high pressure, were chosen through all of the S scenarios but, again, always by controllers working ZKC airspace. Although the traffic levels and complexity in the ZKC sectors indicate that it is likely that scenario load is the reason for the higher ratings, there is a confound in the data. Only active controller participants worked these sectors, making it a possibility that there was a difference in the way active and retired controllers rated their experiences.
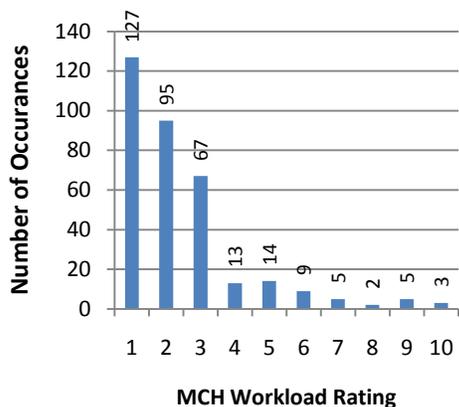

Figure 20.   Distribution of controllers' TLX workload ratings for the M scenarios.
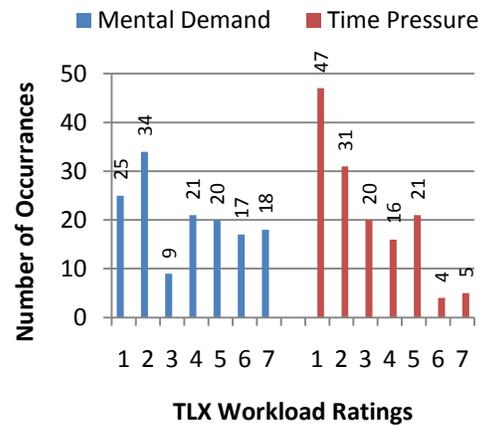

Figure 21.   Distribution of controllers' TLX workload ratings for the S scenarios.

Distributions of the pilots' MCH ratings associated with the M and S scenarios are shown in Figs. 22 and 23. In the airborne experiment, 85 percent of the pilots' MCH workload ratings provided in conjunction with the airborne concept's M scenarios consisted of ratings of "3" or less, and 90 percent of the ratings for the airborne concept's S scenarios consisted of ratings of "3" or less. Overall, these results indicate that the workload experienced by the pilot participants was deemed to be acceptable. However, specific events will be analyzed and described in subsequent papers to elucidate why, at times, pilots provided ratings associated with unacceptable levels of workload. For example, the 15 ratings of "7" or higher recorded in conjunction with the M scenarios, 11 of which were associated with runs simulating 2.0x traffic density, and the four ratings of "7" and "8" recorded in conjunction with the S scenarios, three of which were associated with runs simulating synchronous trajectory change events, will be closely examined.

Figure 22. Distribution of pilots' MCH workload ratings for the M scenarios.
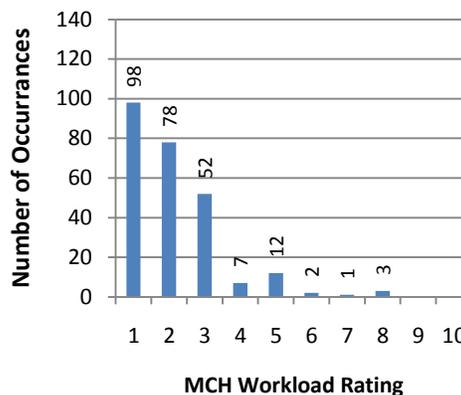
Figure 23. Distribution of pilots' MCH workload ratings for the S scenarios.

Effect of Traffic Density and Schedule Assignment (M Scenarios)

Descriptive statistics for the TLX workload ratings associated with the ground-based concept's M scenarios are shown in Table 17, and descriptive statistics for the MCH workload ratings associated with the airborne concept's M scenarios are shown in Table 18. Statistical analysis of each experiment's respective set of data was performed using nonparametric within-subject tests appropriate for analyzing related samples of ordinal data (i.e., Friedman and/or Wilcoxon tests). Details regarding the airborne concept experiment's workload data analysis are presented elsewhere,[26] and details regarding the ground-based concept experiment's data analysis will be available in subsequent publications. Results associated with each experiment's workload rating data are briefly described below.

For the ground-based experiment, no statistically significant difference was found to exist between the mental demand ratings controllers provided for scenarios involving scheduling assignments (M3 and M4) as compared to scenarios involving no scheduling assignment (baseline scenarios M1 and M2) ($p = 0.327$). Similarly, no statistically significant difference was found to exist between the time pressure ratings controllers provided for scenarios involving scheduling assignments (M3 and M4) versus the baseline scenarios (M1 and M2) ($p = 0.44$). While metering did not have a significant effect on controllers' ratings of their mental and time load, their average ratings under conditions involving scheduling assignments were generally higher. This supports the use of scheduling tools under these conditions because metering adds constraints to the controller tasks, making them more difficult. Having no significant difference between the baseline and metering conditions suggests the tools offset a portion of the additional load created by the metering task.

A statistically significant difference existed between controllers' mental demand ratings of scenarios involving the 1.5x traffic density level (M1 and M4) and scenarios involving the 2.0x traffic density level (M2 and M3) ($p = 0.012$). Furthermore, a statistically significant difference existed between controllers' time pressure ratings of scenarios involving the 1.5x traffic density level (M1 and M4) and scenarios involving the 2.0x traffic density level

Table 17. Descriptive statistics for controllers' TLX workload ratings in the ground-based experiment's M scenarios.

Mental Demand					
	Traffic Density	Schedule Assignment	N	Mean	SD
M1	1.5x	No	16	2.12	0.81
M2	2.0x	No	24	3.50	1.50
M4	1.5x	Yes	16	2.68	1.01
M3	2.0x	Yes	16	3.62	1.65
Time Pressure					
	Traffic Density	Schedule Assignment	N	Mean	SD
M1	1.5x	No	16	1.43	0.73
M2	2.0x	No	24	2.57	1.22
M4	1.5x	Yes	16	1.68	0.79
M3	2.0x	Yes	16	2.81	1.19

Table 18. Descriptive statistics for pilots' MCH workload ratings in the airborne experiment's M scenarios.

	Traffic Density	Schedule Assignment	N	Mean	SD
M1	1.5x	No	85	2.18	1.83
M2	2.0x	No	85	2.33	1.65
M4	1.5x	Yes	85	2.41	1.41
M3	2.0x	Yes	85	2.73	2.18

(M2 and M3) ($p = 0.012$). Thus, greater amounts of traffic and the increase in problem complexity associated with this did have an impact on the controllers as reflected in the level of mental and time load they felt completing the problems.

For the airborne experiment, a statistically significant difference existed between pilots' workload ratings for scenario M1 as compared to M3 ($p = 0.034$). Pilots found that the combination of higher traffic density and an RTA constraint increased workload, whereas either effect separately did not. (Note that pilots were not told the traffic density of each scenario or that the density was changing between scenarios.) No statistically significant difference was found to exist between pilots' workload ratings of scenarios involving the 1.5x traffic density level (M1 and M4) and scenarios involving the 2.0x traffic density level (M2 and M3) ($p = 0.267$). Pilots provided a statistically significant lower mean workload rating for scenarios flown without RTAs (M1 and M2) when compared with scenarios flown with RTAs (M3 and M4) ($p = 0.031$). This RTA effect may be related to the extra effort required to comprehend the RTA data link message, load the RTA in the FMS, and execute the change.

Effect of Timing of the Trajectory Change Event (S Scenarios)

Descriptive statistics for the TLX workload ratings associated with the ground-based concept's three S scenarios are shown in Table 19, and descriptive statistics for the MCH workload ratings associated with the airborne concept's three S scenarios are shown in Table 20. As with the M scenarios, statistical analysis of each experiment's respective data set was performed using nonparametric within-subject tests appropriate for analyzing related samples of ordinal data. A brief description of results is provided below.

For both mental demand and time pressure, controllers reported experiencing lower average workload levels during the basic STA condition (S1) than during either the dispersed or synchronous metering conditions (S2 or S3). The controllers provided fewer of the "very high" mental demand ratings for S1 than for S2 or S3, but no statistically significant difference existed among the mental demand ratings across the three conditions. Additionally, there was no significant difference between the controllers' ratings of the time pressure they felt during each condition, although, again, they used fewer of the "very high" time pressure ratings in S1 than in S2 or S3 and reported a lower average mental workload in conjunction with S1. Thus, in terms of perceived mental and time load, although it made a difference to the distribution of controllers' ratings, the style of metering did not significantly affect their perceptions of workload.

Table 19. Descriptive statistics for controllers' TLX workload ratings in the ground-based experiment's S scenarios.

	Mental Demand			
	Rescheduling	N	Mean	SD
S1	None	48	3.54	1.82
S2	Dispersed	48	3.72	2.18
S3	Synchronous	48	3.81	2.18
	Time Pressure			
	Rescheduling	N	Mean	SD
S1	None	48	2.47	1.59
S2	Dispersed	48	2.83	1.75
S3	Synchronous	48	2.95	1.82

Pilots provided a statistically significant lower mean workload rating for the baseline scenario (S1) than for either scenario involving RTA changes [i.e., scenario S2, which involved an RTA change sent to one aircraft every 10 seconds, ($p = 0.002$); or scenario S3, which involved sending RTA changes to all aircraft at the same time, ($p = 0.009$)]. However, no statistically significant difference was indicated between the ratings provided for scenarios S2 and S3 ($p = 0.982$). Therefore, the inclusion of a second RTA event affected pilot workload, but the relative timing of this event among the 75 aircraft did not affect pilot workload.

2. Ratings of the Operational Concepts and Procedures

In the ground-based concept experiment, controller participants completed a modified version of the Controller Acceptance Rating Scale (CARS)[27] after every run. The scale was broken down into six questions, three of which had a yes/no answer format and three that were answered using three-point rating scales. These latter rating responses and the first yes/no question were combined to create the 10-point rating scale of the CARS, with "1" indicating that the operation was unsafe and unworkable and "10" indicating that the operation was completely acceptable.

Table 20. Descriptive statistics for pilots' MCH workload ratings in the airborne experiment's S scenarios.

	Rescheduling	N	Mean	SD
S1	None	84	1.83	1.16
S2	Dispersed	85	2.30	1.27
S3	Synchronous	84	2.27	1.50

Pilots provided feedback regarding the acceptability of the airborne self-separation operational concept by completing post-scenario and post-experiment questionnaires and participating in post-experiment group debriefing sessions. Questionnaire items were answered using a yes/no format, seven-point rating scales, or text entry fields. As reported in Wing et al.[26], group-debrief sessions involved discussion of several common themes regarding the operational concept in general and its implementation within the experiment environment.

Controller and Pilot Post-Run Assessments of the M scenarios

The controllers' modal acceptability (CARS) rating for the M scenarios was "8," indicating that "minimal controller compensation [was] needed to reach desired performance." Fifty-four percent of the controllers gave the operation this rating, indicating that the operation was satisfactory and had adequate performance. When tested against the four M scenarios (M1 – M4), these CARS ratings were not significantly different at the $p \leq 0.05$ level.

Table 21 splits out the three selections made by controllers to create the CARS: safety, adequacy, and "satisfactoriness." For example, column two of the table shows that 78 percent of responses indicated that the operation was safe. Note, however, that these confirmations were more common in the lower (1.5x) traffic density conditions (M1 and M4), where 91 percent of responses were confirmations, than in the higher (2.0x) traffic density conditions (M2 and M3), where only 67.5 percent of responses were confirmations. This finding suggests that controllers judged the operation performed during the 2.0x traffic density level to be safe only two-thirds of the time, whereas they judged the operation to be safe nine-tenths of the time when performed during the 1.5x traffic density level. This lower safety assessment in the 2.0x scenario is consistent with the ground-based LOS discussion included in the safety section of this paper, as the departure flows with the highest complexity leading to separation problems occurred in the 2.0x scenarios.

Table 21. Confirmation counts for the ground-based concept over the four M scenarios' CARS ratings.

	Traffic Density	Schedule Assignment	Operation Rated "Safe"	Operation Rated "Adequate"	Operation Rated "Satisfactory"	Total Ratings Above "Acceptable"
M1	1.5x	No	*15 (of 16)*	*14 (of 15)*	*14 (of 14)*	*14*
M2	2.0x	No	*17 (of 24)*	*17 (of 17)*	*16 (of 17)*	*16*
M4	1.5x	Yes	*14 (of 16)*	*14 (of 14)*	*14 (of 14)*	*14*
M3	2.0x	Yes	*10 (of 16)*	*10 (of 10)*	*9 (of 10)*	*9*
			Total = 56 (of 72), or 78%	*Total = 55 (of 56), or 98%*	*Total = 53 (of 55), or 96%*	*Total = 53 (of 72), or 73.6%*

A Friedman test indicated that there was not a significant difference in controllers' ratings of safety by scenario. Separate Wilcoxon tests of the traffic and metering conditions showed that the level of traffic, although it accounts for the difference, was not significant.

Pilots' post-scenario questionnaire responses indicate that the majority of the airborne concept experiment's M scenarios were perceived as involving safe operations. When asked if they felt that safety was ever compromised, pilots responded "yes" after 39 of the M scenarios and responded "no" after 325[*] of the M scenarios; i.e., pilots reported experiencing unsafe conditions during approximately 11 percent of the M scenarios. The majority (67 percent) of the unsafe conditions were reported to be experienced during scenarios involving the higher (2.0x) traffic density level (28 percent during M2, and 38 percent during M3). In some instances, pilots commented that the single-pilot testing of the airborne concept in this experiment precluded the opportunity for crew cross-checks of conflict resolutions and other trajectory changes, which would be paramount for safety in an operational setting.

Controller and Pilot Post-Run Assessments of the S scenarios

As with the M scenarios, the controllers' modal CARS rating for the S scenarios was "8," indicating that "minimal controller compensation [was] needed to reach desired performance." Sixty-three percent of the controllers gave the operation this rating, indicating that they felt it was satisfactory and had adequate performance. When tested against the three rescheduling conditions (S1, S2, and S3), using a Friedman test, these CARS ratings were not significantly different.

Table 22 splits out the three selections made by controllers to create the CARS – based on safety, adequacy, and "satisfactoriness." For example, column two of the table shows that 79 percent of responses said the operation was safe, and these confirmations were fairly evenly spread across the three conditions. The table shows that approximately the same number of responses associated with each type of metering judged the ground-based operation to be firstly safe and then adequate (approximately 75 to 80 percent). However, when considering the non-rescheduled condition (S1), an eighth of the controllers' responses to the satisfactory question were negative –

---

[*] A total of 368 M scenarios were completed by the airborne concept experiment's pilot participants (i.e., eight scenarios were completed by each of 46 pilots). However, only 364 responses were provided for this particular post-scenario questionnaire item.

Table 22. Confirmation counts for the ground-based concept over the three S scenarios' CARS ratings.

	Rescheduling	Operation Rated "Safe"	Operation Rated "Adequate"	Operation Rated "Satisfactory"	Total Ratings Above "Acceptable"
S1	None	*37 (of 48)*	*37 (of 37)*	*31 (of 37)*	*31*
S2	Dispersed	*39 (of 48)*	*38 (of 39)*	*36 (of 38)*	*36*
S3	Synchronous	*38 (of 48)*	*36 (of 38)*	*36 (of 36)*	*36*
		Total = 114 (of 144), or 79%	*Total = 111 (of 114), or 97%*	*Total = 103 (of 111), or 93%*	*Total = 103 (out of 144), or 71.5%*

enough to make this condition stand out as less acceptable than the other two conditions, even though S1 ratings were not found to be significantly different. This fact indicates that controllers found a non-adjusting schedule less acceptable than one that adjusted, but when rescheduling occurred, the two timing schemes were equally acceptable.

Pilots' post-scenario questionnaire responses indicate that the majority of the airborne concept experiment's S scenarios were perceived as involving safe operations. When asked if they felt that safety was ever compromised, pilots responded "yes" after 16 of the S scenarios and responded "no" after 274[*] of the S scenarios; i.e., pilots reported experiencing unsafe conditions during approximately 6 percent of the S scenarios. The majority (81 percent) of the unsafe conditions were reported to be experienced during scenarios involving trajectory change events (38 percent during S2, and 43 percent during S3).

Controller and Pilot Post-Experiment Assessments

At the end of the experiment, the ground-based study's 18 controller participants were asked to rate the acceptability of their role in the ground-based concept. They provided a mean response of 4.33 ("acceptable") on a five-point scale (SD = 0.68, N = 18), where "1" corresponded to "completely unacceptable," and "5" corresponded to "completely acceptable." They also thought that the ground-based automation offered "some timely support" (M=3.88 on a 5-point, low to high scale, SD=0.58, N=18) when asked if the automation supported them at the level they needed during the study.

One of the main differences in the ground based study between the concept and the current day is that the automation makes an assessment before it presents information to the controllers. The controllers were asked if they felt the automation left them enough time to work on the problems it had identified. They responded that the automation gave them a "reasonable" amount of time to work on problems (M=3.44 on a 5-point, low to high scale, SD =1.24, N=18).

For the ground-based study, the issue of needing more assistance with increased task-load was addressed by asking controllers whether they had to rely more on the automation tools as traffic level and complexity increased. Controllers said they "relied on the tools a lot more" (M=4.88 on a 5-point, low to high scale, SD =0.48, N=18). Controllers also rated how comfortable they were with their task load at the end of each run. In both the M scenarios and the S scenarios, controllers' mean responses were that their task load was "light" (for M scenarios: M=2.64, SD = 1.52, N = 72; for S scenarios: M=3.11, SD = 1.73, N = 144), when using a scale of 1 ("too few tasks") to 7 ("too many tasks").

Pilots' post-experiment questionnaire responses indicate that they felt they could use airborne self-separation in a relatively effective manner and that they found the operational concept to be somewhat acceptable given their relatively limited exposure to the concept's implementation within one experiment environment. Using a scale of 1 ("very effectively") to 7 ("very ineffectively") to rate how effectively they could use airborne self-separation to operate within high density en-route airspace, the pilots' mean response was 2.41 (SD = 1.02, N = 46). Similarly, when asked to rate the acceptability of the airborne self-separation concept using a scale of 1 ("very acceptable") to 7 ("very unacceptable"), the mean response was 2.54 (SD = 1.26, N = 46).

Pilots reported that they would be somewhat comfortable using the operational concept to fly through high density en-route airspace, and they indicated that they felt the safety of airborne self-separation, as it was represented in the airborne concept experiment, would have a safety level comparable to that associated with current day operations. The pilots provided a mean response of 3.02 (SD = 1.31, N = 46) when using a scale of 1 ("very comfortable") to 7 ("very uncomfortable") to rate how comfortable they would be flying through high density en-

---

[*] A total of 276 S scenarios were completed by the airborne concept experiment's pilot participants (i.e., six scenarios were completed by each of 46 pilots). However, only 274 responses were provided for this particular post-scenario questionnaire item.

route airspace requiring the use of airborne self-separation, and they provided a mean response of 3.65 (SD = 1.29, N = 46) when using a scale of 1 ("safety is enhanced") to 7 ("safety is compromised") to describe their overall assessment of the safety of airborne self-separation. During the group debrief sessions, pilots commented that their perception of safety was reduced by occasions of short-notice conflict alerts and situations where the automation provided no resolutions.

With respect to workload, 80 percent of the airborne concept experiment's pilot participants reported that the self-separation procedures they were asked to use would serve as an acceptable workload trade-off when compared with current day procedures for maintaining separation. When asked to rate the acceptability of the new self-separation procedures in terms of required workload level, when compared with current day procedures, the pilots provided a mean response of 3.13 (SD = 1.39, N = 46) using a scale of 1 ("very acceptable workload level") to 7 ("very unacceptable workload level"). When asked to rate how well they thought the self-separation procedures could be integrated into the current flight deck operational environment, assuming the availability of appropriate decision support tools, the pilots provided a mean response of 2.67 (SD = 1.15, N = 46) using a scale of 1 ("can be integrated easily") to 7 ("cannot be integrated").

VI. Summary of Findings

In a coordinated effort, two human-in-the-loop simulation experiments of two operational concepts were jointly designed and conducted in parallel to explore function allocation in separation assurance. In the centralized concept *ground-based automated separation assurance* ('ground-based concept'), ground-based automation predicts aircraft trajectories, detects conflicts, computes resolution trajectories, and issues trajectory amendments automatically to aircraft by data link communication if the resolution is within acceptable limits. The air traffic controller monitors the operation and is available for providing services and handling exceptions, such as creating or approving trajectories when the initial automated solution is outside the predefined tolerances. The pilot, in a role unchanged from current day operations, executes the instructions in a timely manner. In the distributed concept *airborne trajectory management with self-separation* ('airborne concept'), airborne automation predicts aircraft trajectories, detects conflicts, alerts the pilot, computes resolution trajectory alternatives, and displays these alternatives to the pilot. The pilot selects from among the alternative resolutions and executes the new trajectory. The controller performs no separation function for the equipped aircraft, but the Air Navigation Service Provider supplies traffic flow management constraints to equipped aircraft. The two experiments tested homogeneous operations in nominal conditions, i.e., all aircraft operating under the same separation scheme and no failure modes or other conditions disruptive of normal procedures.

The principal goal of these coordinated experiments was to determine how to achieve comparability of results from different simulation platforms while providing baseline results for future experiments. Plans for future studies include introducing mixed operations and off-nominal conditions. The series of experiments together are intended to illuminate the performance and other attributes of function allocation for separation assurance across a range of conditions. The approach taken to achieve comparability was to run common traffic scenarios in a set of common test matrices. The independent variables were traffic density (1.5x and 2.0x of current day levels), arrival scheduling (without and with), and the timing of arrival rescheduling (none, dispersed assignment times, synchronous assignment times). Common metrics were defined to aid in comparability of results. The joint experiments achieved the goal of producing comparable results in some metrics, but not all metrics were suitable for direct comparison.

1. Metrics Suited for Direct Comparability

Comparability was maximized in metrics primarily associated with automation performance: conflicts detected, times before predicted separation loss of initial detection and final resolution, flight path efficiency, and arrival time conformance. Performance in these respects was generally similar and acceptable, with most differences attributable to differing automation systems designs.

<u>Conflict Detection and Resolution</u>

The airborne system was observed to generally detect more conflicts than the ground-based system and with less consistent behavior with regard to traffic density and schedule constraints. A preliminary analysis indicated that the difference may be attributable in part to the vertical buffers applied by the airborne system to transition segments (climbs or descents), causing significantly more conflicts to be detected. The ground-based approach appeared to show greater sensitivity to the traffic density manipulation regarding conflicts detected and times of initial detection and final resolution than the airborne approach, but its conflicts were generally detected and resolved sooner. In both experiments, conflicts were cleared with greater than five minutes on average remaining to the predicted

separation loss for both concepts, indicating acceptable mean conflict resolution performance. Nevertheless, the number of occurrences of conflicts detected with much shorter time remaining was a significant safety concern in both concepts that will require improvement. More analysis is underway to understand all the factors that contributed to unexpected conflict detection and resolution behavior.

Efficiency

The efficiency metrics of flight path deviation and arrival schedule conformance indicated similar results for the two concepts. Mean flight path deviation increased with traffic density and was not affected by the inclusion of arrival time constraints. An arrival schedule change increased the mean flight path length (as expected of delay maneuvers) in the airborne experiment but not in the ground-based experiment. The difference was attributed to the ground controllers not having adequate time to complete the assignment of delay maneuvers to all arriving aircraft prior to scenario termination and thus fewer aircraft having a changed path length as compared to the same aircraft in the airborne experiment.

With the change in arrival schedule, the aircraft in the airborne experiment had better mean arrival time conformance and less dispersion but equivalently excellent median conformance as compared to aircraft in the ground-based experiment. Mean arrival time conformance was better without rescheduling than with rescheduling for both concepts, attributed in the ground-based experiment to controllers having inadequate time within the short scenario to reschedule flights and in the airborne experiment to some pilots electing to postpone the delay maneuver execution until later in the flight. Both results were artifacts of the limited scenario duration in the simulations. The relative timing of the rescheduling assignments (dispersed or synchronous among aircraft) had no significant effect in the efficiency metrics of either experiment.

2. Metrics Not Directly Comparable

Direct comparability was not achievable for some metrics, although limited comparisons were possible in some areas and some similar themes emerged in others.

Resolution Type

The type of conflict resolutions executed (strategic or tactical) had only limited comparability because of significant differences in system design. Both concepts had tactical modes that were activated in situations with only a short time available for resolution, but this time was one to two minutes earlier for the airborne concept. Thus, the airborne experiment registered more occurrences of tactical resolutions than did the ground-based experiment.

Loss of Separation

Losses of separation were experienced in both experiments, but the quantity and nature of the events were unique to each simulation and cannot be compared. In the ground-based experiment, late detections and complex interactions of traffic flows that were used as background traffic in the airborne experiment were the primary contributing factors. In the airborne experiment, simulation faults and pilot procedural errors were the primary contributing factors. For both concepts, analysis of the separation loss data will result in system and procedural improvements to reduce the likelihood of reoccurrence of these failure modes in future studies.

Workload and Concept Acceptability

Subjective assessments by the subject participants could not be directly compared because of their different roles and perspectives in their respective concepts. In assessing workload, the controllers in the ground-based experiment indicated that average to below average mental demand was required, and they felt below average to low time pressure to complete their tasks. The pilots in the airborne experiment indicated a difficulty level of fair or mild and that an acceptable level of mental effort was required to attain adequate performance. Specific instances of high subjective workload were registered in both experiments, and these instances will be further analyzed to determine what system and procedural improvements will be required. In the ground-based experiment, workload ratings increased with traffic density but not with the inclusion of schedule constraints. In the airborne experiment, workload ratings increased with the inclusion of scheduling constraints but not with increased traffic density. In both experiments, rescheduling increased workload but the relative timing of the rescheduling among aircraft had no effect.

In assessing concept acceptability, the controllers in the ground-based experiment and the pilots in the airborne experiment felt the concepts were safe for the majority of runs but not for all runs. Debrief sessions indicated that short-term conflict alerts were a significant contributor to the perception that some scenarios were unsafe. As noted earlier, causes of short-term conflict alerts included operational factors such as the difficulty of accurately predicting conflicts on vertical segments and simulation factors such as software errors that resulted in the broadcast of incorrect information over ADS-B. Addressing these factors will be a priority in the preparation of future experiments.

Overall, controllers found their role in the ground-based concept to be acceptable. Pilots in the airborne experiment indicated they would feel somewhat comfortable using airborne self-separation in high density en-route airspace and that the safety level would be comparable to that associated with current day operations.

VII. Conclusions and Recommendations

This paper describes a first step within NASA's multi-year research plan to study advanced function allocation concepts for NextGen separation assurance in high density airspace. This first pair of experiments addressed, homogenous operations under nominal conditions. Future experiments will address mixed operations and off-nominal conditions. At the end of this research, we hope to have identified and quantified the most viable options for allocating the functions between the air and the ground, as well as the human and the automation, and to be able to provide concrete recommendations for implementation from a technical perspective. This results of the first experiments described in this paper are not intended to provide such guidance. These experiments were designed to initiate the process of providing directly comparable results when conducting research on different function allocation concepts in different laboratories, while also gathering additional baseline data about the operational concepts under investigation to provide the foundation for the future planned studies.

Providing comparable results is a very challenging task, and while many efforts were made, only small subsets of the data from this experiment could actually be directly compared. Nevertheless, where comparisons were possible, no substantial differences in performance or operator acceptability were observed. Mean schedule conformance and flight path deviation were considered adequate for both approaches. Conflict detection warning times and resolution times were mostly adequate, but certain conflict situations were detected too late to be resolved in a timely manner. This led to some situations in which safety was compromised and/or workload was rated as being unacceptable in both experiments. Operators acknowledged these issues in their responses and ratings but gave generally positive assessments of the respective concept and operations they experienced. Future studies will evaluate technical improvements and procedural enhancements to achieve the required level of safety and acceptability.

The function allocation concept between the operator and the automation that is pursued within each individual approach, airborne and ground-based, appears to be tangible and on the right track but needs further refinement. Gaining additional insight into the function allocation between the air and the ground is planned as the next step. With these homogeneous performance baseline studies now completed, future studies will investigate the integration of airborne and ground-based capabilities within the same airspace to leverage the benefits of each concept.

As these future experiments will have to provide data that enable a true comparison of performance and problems with different approaches, the intended comparisons have to be further integrated into the experimental design. While several factors were included in this first experiment, this activity also highlighted areas that could be improved to increase the range of comparable results.

References

[1]Pearce, R.A. The NextGen JPDO Model of Interagency Planning. AIAA 2009-7010, 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO), Hilton Head, South Carolina, 2009.

[2]Federal Aviation Administration. FAA Aerospace Forecast Fiscal Years 2010-2030. Available at http://www.faa.gov/data_research/aviation/aerospace_forecasts/2010-2030/media/2010 Forecast Doc.pdf

[3]Kopardekar, P., Rhoades, J., Schwartz, A., Magyarits, S., and Willems, B. Relationship of Maximum Manageable Air Traffic Control Complexity and Sector Capacity. 26th International Congress of the Aeronautical Sciences. Anchorage, AK. 2008.

[4]Prevot, T., Homola, J., Mercer, J., Mainini, M. and Cabrall, C. "Initial Evaluation of NextGen Air/Ground Operations with Ground-Based Automated Separation Assurance," Eighth USA/Europe Air Traffic Management Research and Development Seminar, Napa, California, 2009.

[5]Wing, D. A Potentially Useful Role for Airborne Separation in 4D-Trajectory ATM Operations. AIAA 2005-7336, 5th AIAA Aviation Technology, Integration, and Operations Conference (ATIO), 26-28 September 2005, Arlington, Virginia.

[6]FAA (2010) NAS Enterprise Architecture Portal https://nasea.faa.gov

[7]Johnson W, Battiste V, Delzell S, Holland S, Belcher S and Jordan K. Development and demonstration of a prototype free flight CDTI, SAE/AIAA World Aviation Congress, Anaheim, CA, 1997.

[8]Erzberger H. Transforming the NAS: The Next Generation Air Traffic Control System, 24th International Congress of the Aeronautical Sciences, Yokohama, Japan, 2005.

[9]Prevot T, Smith N, Palmer E, Mercer J, Lee P, Homola J, and Callantine T. The Airspace Operations Laboratory (AOL) at NASA Ames Research Center, AIAA-2006-6112, AIAA, Reston, VA, 2006.

[10]Prevot T, Lee P, Callantine T, Mercer J, Homola J, Smith N, et al. (in press). Human-in-the-Loop Evaluation of NextGen Concepts in the Airspace Operations Laboratory. *Proceedings of the AIAA Modeling and Simulation Technologies (MST) Conference,* Toronto, Canada, 2010.

[11]Prevot T. Exploring the Many Perspectives of Distributed Air Traffic Management: The Multi-Aircraft Control System MACS. In S. Chatty, J. Hansman, & G. Boy. (Eds). HCI-Aero 2002, AIAA Press, Menlo Park, CA. pp 149-154, 2002.

[12]Erzberger H, Lauderdale T, and Chu Y-C. Automated Conflict Resolution, Arrival Management and Weather Avoidance for ATM, Proc 27th International Congress of the Aeronautical Sciences, Nice, France, 2010.

[13]Prevot T. NextGen Technologies for Mid-Term and Far-Term Air Traffic Control Operations. 28th Digital Avionics Systems Conference, Orlando FL, 2009

[14]Erzberger H and Heere K. Algorithm and operational concept for solving short range conflicts. International Congress of the Aeronautical Sciences 2008-8.7.5, Anchorage, Alaska, 2008.

[15]Finkelsztein D, Lung T, Vivona R, Bunnell J, Mielke D and Chung W. Airspace and Traffic Operations Simulation for Distributed ATM Research and Development. AIAA Modeling and Simulation Technologies Conference, AIAA 2005-6488. San Francisco, CA, 2005.

[16]Bussink F, Hoekstra J and Heesbeen B. Traffic Manager: A Flexible Desktop Simulation Tool Enabling Future ATM Research. 24th Digital Avionics Systems Conference, ISSN 0-7803-9307-4, Washington D.C., October – November 2005.

[17]RTCA, Inc. Minimum Aviation System Performance Standards for Automatic Dependent Surveillance Broadcast (ADS‑B). DO-242A, RTCA, Washington D.C., 2002.

[18]Palmer M and Ballin G. A High-Performance Simulated On-Board Avionics Architecture to Support Traffic Operations Research. AIAA Modeling and Simulation Technologies Conference, AIAA-2003-5452. Austin, TX, 2003.

[19]Karr D, Roscoe D and Vivona R. An Integrated Flight-Deck Decision-Support Tool in an Autonomous Flight Simulation. AIAA Modeling and Simulation Technologies Conference, AIAA-2004-5261, Providence, RI, 2004.

[20]Karr D, Roscoe D and Vivona R. Conflict Detection Using Variable 4D Uncertainty Bounds to Control Missed Alerts. AIAA Guidance, Navigation, and Control Conference, AIAA-2006-6057, Keystone, CO, 2006.

[21]Vivona R, Karr D and Roscoe D. Pattern-Based Genetic Algorithm for Airborne Conflict Resolution. AIAA Guidance, Navigation, and Control Conference, AIAA-2006-6060, Keystone, CO, 2006.

[22]Wing D, Roscoe D and Vivona R. Airborne Tactical Intent-Based Conflict Resolution Capability. 9th AIAA Aviation Technology, Integration, and Operations Conference, AIAA-2009-7020, Hilton Head, SC, 2009.

[23]Prevot T, Mercer J, Homola J, Cabrall C, and Brasil C. Functional Allocation for Ground-Based Automated Separation Assurance in NextGen. To appear at HCI Aero Conference, Cape Canaveral, FL., Nov 2010.

[24]Hart S and Staveland L. Development of the National Aeronautics and Space Administration (NASA) Task Load Index (TLX): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds), *Human mental workload* (pp139-183). Amsterdam, North Holland, 1988.

[25]Wierwille W and Casali J. A valid rating scale for global mental workload measurement. Proceedings of the Human Factors Society 27th Annual Meeting, Norfolk, VA, October 1983.

[26]Wing D, Murdoch J, Chamberlain J, Consiglio M, Hoadley S, Hubbs C, and Palmer M. Function Allocation with Airborne Self-Separation Evaluated in a Piloted Simulation. Proc 27th International Congress of the Aeronautical Sciences, Nice, France, 2010.

[27]Lee K, Kerns K, Bone R, and Nickelson M. Development and validation of the Controller Acceptance Rating Scale (CARS): Results of empirical research. *4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, NM, December 2001.