Evaluating NextGen Closely Space Parallel Operations Concepts with Validated Human Performance Models

Brian F. Gore[1], Becky L. Hooey[1], Connie M. Socash[2], Nancy Johnson Haan[3], Eric Mahlstedt[1], Deborah L. Bakowski[1], Andrew M. Gacy[2], Christopher D. Wickens[2], Mala Gosakan[2], and David C. Foyle[4]

NASA Ames Research Center
Human Systems Integration Division
MS 262-4, PO Box 1
Moffett Field, CA 94035-0001

Human Centered Systems Laboratory (HCSL) Technical Report (HCSL-11-01)

in support of
FAA/NASA Agreement DTFAWA-10-X-80005
Annex 1.5 -- Closely Spaced Parallel Operations
Task Number: 03-03; 09-AJP61FGI-0052
Deliverables #1 and #2
January 31, 2011

APPROVED FOR PUBLIC DISTRIBUTION

---

[1] San Jose State University
[2] Alion Science and Technology Micro Analysis and Design Operation
[3] Dell Services Federal Government
[4] NASA Ames Research Center

Executive Summary

The objectives of the current research are to: develop valid human performance models (HPMs) of approach and land operations; use these models to evaluate the impact of NextGen Closely Spaced Parallel Operations (CSPO) on pilot performance; and draw conclusions regarding flight deck displays and pilot roles and responsibilities for NextGen CSPO concepts. This research represents the results of the first, of a two-year effort, in which the specific goals were to develop and validate baseline HPMs of current-day RNAV and NextGen CSPO approaches and conduct 'what-if" manipulations of off-nominal scenarios.

Using NASA's Man Machine Integration Design and Analysis System v5 (MIDAS v5), a high-fidelity model of a two-pilot commercial crew flying current-day area navigation (RNAV) approach and land operations was developed. The model, containing over 970 individual pilot tasks, was based on cognitive task analyses and cognitive walkthroughs conducted with commercial pilots and air traffic controllers.

The current-day RNAV model was validated using a methodical, multi-dimensional approach. The model inputs, including the task trace and input parameters, were validated using focus group sessions comprised of a total of 8 commercial pilots with glass-cockpit aircraft and RNAV flying experience. The pilot-centric scenario-based cognitive walkthrough approach captured the context of operations from 10,000' to Touchdown and enabled pilots to assess the modeled tasks and identify tasks that were missing, or in the wrong sequence. The pilots also completed quantitative rating scales, which were used to validate the model input parameters for workload and visual attention. The model was refined based on the results of this input validation process. Next the model outputs, workload and visual attention, of the refined model were statistically compared to existing human-in-the-loop data. The workload model output correlated with a comparable Human-in-the-Loop (HITL) study with r^2 of .54 for overall workload. The individual workload dimensions also correlated positively with the HITL study with r^2 ranging from .55 to .94. Visual percent dwell time correlated with three HITL studies (r^2 = .99). These validation results provide confidence that the model validly represents pilot performance.

Next, the validated baseline RNAV model was extended to represent the Very Closely Space Parallel Approach (VCSPA) concept developed and evaluated at NASA Ames Research Center. Two operational implementations of the VCSPA concept were evaluated: 1) VCSPA 800', with a 800' ceiling and manual flight after a DH of 650' and 2) VCSPA 200' with a 200' ceiling and autoland capability. The model inputs (task trace and input parameters) were validated using the same scenario-based focus groups as above. The VCSPA models' outputs were compared to the RNAV model output. Compared to current-day RNAV approaches, the VCSPA 800' model predicted that the VCSPA 800' implementation will increase visual, auditory, cognitive, and motor workload during the land phase of flight (650' to touchdown). The VCSPA 200' model

predicted that the VCSPA 200’ implementation may reduce cognitive and motor workload as compared to the RNAV and the VCSPA 800’, due primarily to advanced automation. The MIDAS model predicts that both VCSPA scenarios will draw the pilots’ attention to the traffic and wake information on the Navigation (Nav) Display at the expense of attending to the Primary Flight Display (PFD) and Out-the-window (OTW).

Next, “what-if” scenarios were conducted to explore the impact of the NextGen VCSPA concept on pilot performance during *off-nominal events*. A comprehensive analysis was conducted to identify appropriate off-nominal events including semi-structured interviews with the eight pilots in the focus group session (described earlier), a search of the ASRS database, and an analysis using a previously developed taxonomy of off-nominal events and contributing factors. This analysis yielded 13 potential off-nominal events during approach and land. From these, four off-nominal events were selected: 1) High wind/turbulence, 2) Flight Mode Annunciator (FMA) Error, 3) Required Navigation Performance (RNP) alert and 4) Rogue aircraft on the runway. Model outputs including workload, visual attention, and time to detect or respond to the event were recorded across ten runs of each off-nominal model. Implications for CSPO operations including pilot roles and responsibilities and flight deck displays are discussed.

In summary, a methodical and comprehensive process was undertaken to develop and validate models of current-day RNAV and NextGen VCSPA operations. The models were extended to examine “what-if” off-nominal scenarios. The findings yielded implications for candidate NextGen roles and responsibilities and flight deck displays and automation, which will be evaluated in year two (2011) of this research effort.

Table of Content

I. INTRODUCTION

The National Airspace System (NAS) in the United States is currently being redesigned because it is anticipated that the current air traffic control (ATC) system will not be able to manage the predicted two to three times growth in air traffic in the NAS (JPDO, 2009). The goal of the Next Generation Air Transportation System (NextGen) is to increase the capacity, safety, efficiency, and security of air transportation operations (JPDO, 2009). However, in doing so, it is expected that the data available to pilots on the flight deck (e.g., weather, wake, traffic trajectory projections, etc.) will be increased substantially in order to support more precise and closely coordinated operations. If not designed with consideration of the human operators' limitations and capabilities, these NextGen concepts could leave pilots, and thus the entire aviation system, vulnerable.

Human Performance Models (HPMs) have been shown to play a role in all phases of the concept development, refinement, and deployment process of next generation systems (Gore, 2008; Hooey & Foyle, 2008). HPMs can be used to develop and evaluate new technologies, operational procedures, and the allocation of roles and responsibilities between the human operators and the automation (Gore, 2008; Gore & Corker, 1999; Hooey & Foyle, 2008). HPMs hold the most promise when they are used early in the system design, or system redesign, process and when used iteratively with human-in-the-loop (HITL) simulation output (Elkind, Card, Hochberg, & Huey, 1989; Gore, 2002; Hooey & Foyle, 2008). However, before HPMs can be successfully implemented to evaluate how NextGen concepts will impact pilot performance, baseline models of current-day pilot performance must first be developed and validated.

The objective of this research effort is to develop, validate and extend a baseline HPM of pilot performance to evaluate proposed changes to flight deck technologies and pilot procedures, operations, and roles and responsibilities that will be likely with the NextGen concept termed Closely Spaced Parallel Operations (CSPO). Model outputs, including time required to complete tasks, event/error detection, pilot workload, and visual attention will be used to draw conclusions regarding the requirements necessary to support NextGen concepts and predict the human performance effects, identify safety vulnerabilities, and recommend mitigations.

In Section I of this report, high-level goals for CSPO in the NAS and one CSPO concept, Very Closely Spaced Parallel Approaches will be presented as well as a description of the Man-machine Integration Design and Analysis System, a human performance modeling tool. In Section II, a comprehensive MIDAS model of current-day RNAV approach-and-land operations is presented. The methodical, iterative, approach used for model development and validation will be discussed. In Section III, the validated RNAV model is extended to include two instantiations of the VCSPA concept and the impact of this NextGen concept on pilot performance is discussed. Finally, in Section IV, four 'what-if' off-nominal scenarios are presented with implications for NextGen flight deck displays and pilot roles and responsibilities.

CLOSELY SPACED PARALLEL OPERATIONS

CSPO is a proposed solution to increase efficiency and throughput in low-visibility conditions by enabling parallel approaches to be conducted in instrument meteorological conditions (IMC) with lower landing minima (JPDO, 2009). The current requirement for landings in IMC is at least 4300’ of lateral runway spacing (as close as 3000’ for runways with a Precision Runway Monitor) whereas operations in visual meteorological conditions (VMC) require lateral runway spacing to be equal to or greater than 750’.

It is feasible for aircraft to perform both arrival and departure operations in IMC using VMC parallel separation standards as advanced navigation technology, datalink, sophisticated wake avoidance algorithms, 4-D flight management systems, and advanced flight deck displays become more widely available. As flightdecks are modified to accommodate the new suite of automation tools and displays, research must be conducted to ensure they are designed and implemented in a safe manner without leaving pilots vulnerable to errors or excess workload.

Very Closely Spaced Parallel Approach Scenarios

The Very Closely Spaced Parallel Approach (VCSPA) concept, based on Raytheon’s Terminal Area Capacity Enhancement (TACEC) parallel approach procedures, is intended to increase efficiency and throughput in low-visibility conditions. The concept proposes parallel operations in IMC conditions with lower landing minimas and runway spacing reduced to 750’ (Verma, Lozito, & Trott, 2008). Because this reduction in runway spacing increases the likelihood of wake vortex incursions advanced equipage is assumed including: Global Positioning Sattelite (GPS), Automatic Dependent Surveillance-B (ADS-B), Datalink, 4-D Flight Management System (FMS; including speed-coupling algorithms), and dedicated or augmented displays for wake information. The VCSPA/O concept also requires that a safe breakout maneuver be dynamically calculated and presented real-time to the crew on the flight deck (Verma, et al., 2008), effectively transferring the responsibility for detecting a transgression and initiating the breakout maneuver from ATC to the flightdeck.

In the VCSPA concept, aircraft are paired by ATC approximately 30 nm from the runway threshold. Pairs are based on aircraft performance, arrival direction, and aircraft weight. The trailing aircraft is slewed 6 deg from runway (Figure 1). At approximately 12 to 17 nm from runway threshold, ATC initiates self-separation operations by datalink message, and the pilot engages the speed-coupling automation mode on the mode control panel (MCP). Once this action is taken, the flight mode annunciator (FMA) shows C–SPD, C-LNAV, C-VNAV. Operationally, the trailing aircraft maintains 12 sec spacing behind the lead aircraft using speed-algorithms and 4-D FMS automation. At approximately 2 nm from the runway threshold, or 1,100’, the trailing aircraft then aligns with runway and is parallel with the lead aircraft.

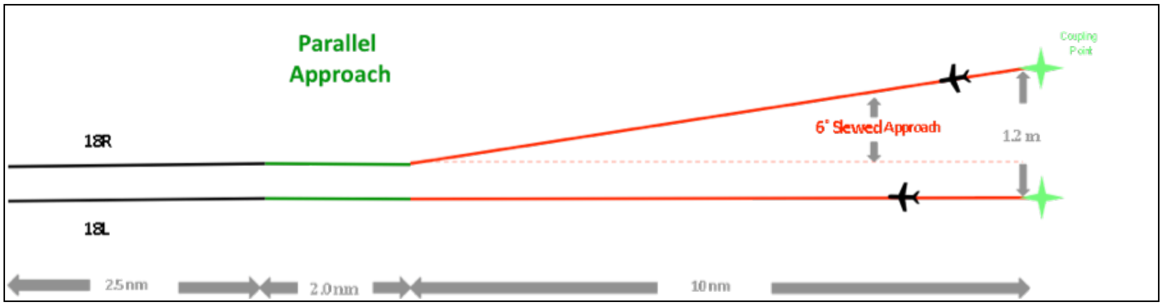


Figure 1. Very Closely Spaced Parallel Approach (VCSPA) Concept.

When new technologies and procedures are introduced into flight deck operations, as are expected with CSPO concepts, such as VCSPA, new human-system vulnerabilities may surface. Human Performance Models (HPMs) may be used to predict the points at which the human-system vulnerabilities are most likely. One such HPM tool, the Man-machine Integration Design and Analysis System (MIDAS), is discussed next.

MIDAS

The Man-machine Integration Design and Analysis System (MIDAS) is an established HPM that predicts human-system performance under nominal and off-nominal conditions (Gore, 2008). MIDAS is a dynamic, integrated human performance modeling environment that facilitates the design, visualization, and computational evaluation of complex man-machine system concepts in simulated operational environments. MIDAS symbolically represents many mechanisms that underlie and cause human behavior. MIDAS combines graphical equipment prototyping, dynamic simulation, and HPMs to reduce design cycle time, support quantitative predictions of human-system effectiveness, and improve the design of crew stations and their associated operating procedures.

Figure 2 illustrates the organization and flow of information among the MIDAS components. For a description of the MIDAS processes, the reader is directed to Gore 2010a.

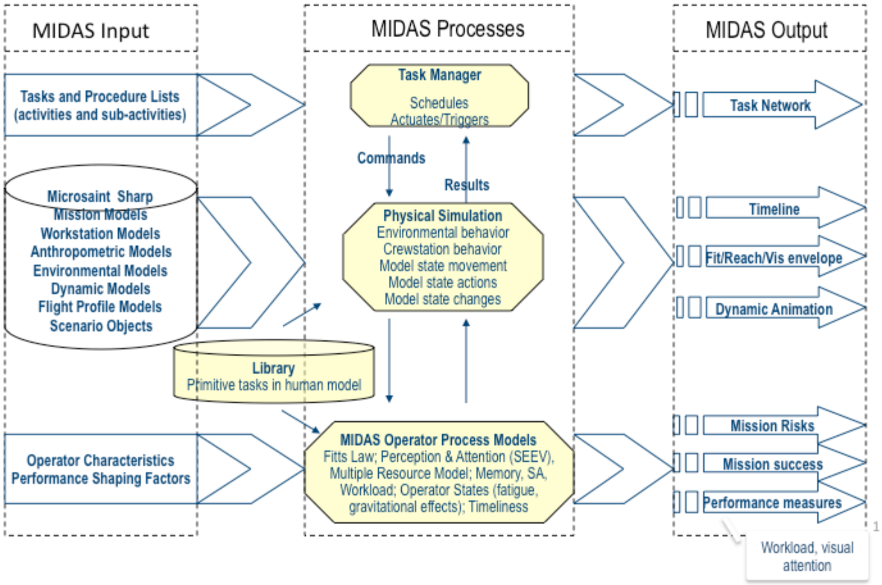


Figure 2. MIDAS Architecture (adapted from (Gore, 2010).

MIDAS inputs (Figure 2, left column) include the operators' task and procedures, the operational environment (e.g., flight profiles, scenario objects and events, cockpit layout etc), and operator characteristics (e.g., algorithms that represent operator expertise, and fatigue).

The **MIDAS processes** (Figure 2, middle column) are comprised of a task manager model that schedules tasks to be completed, definitions of the state of models within the physical simulation, a library of "basic" human primitive models that represent behaviors required for all activities, and process models such as operator perception, visual attention, and workload. Figure 3 illustrates the process models contained with MIDAS and the empirical basis upon which they were generated and validated.

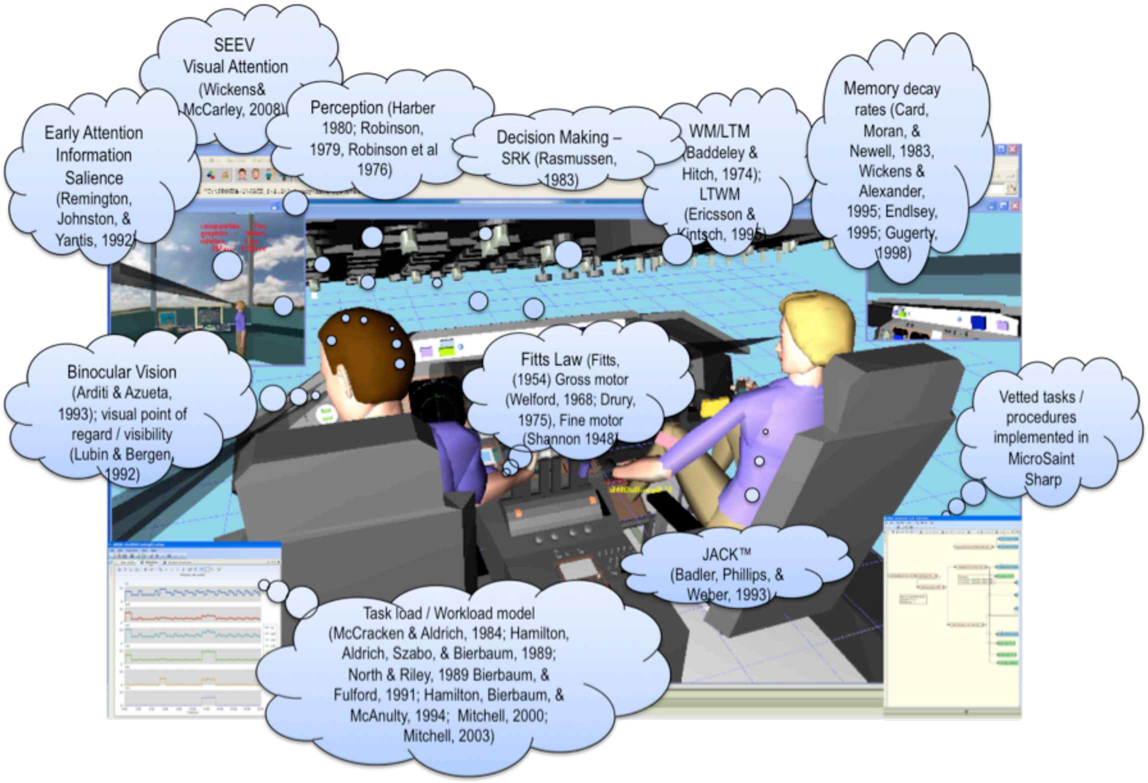


Figure 3. MIDAS v5 Process Models.

These process models have been extensively validated. For instance, MIDAS' attention-guiding model operates according to the SEEV model (Wickens & McCarley, 2008), an extensively validated model that estimates the probability of attending, P(A), to an area of interest in visual space, as a linear weighted combination of the four components - salience, effort, expectancy, and value. Attention in dynamic environments is driven by the bottom-up capture of *Salient (S)* events (e.g., a flashing warning on the instrument panel) and inhibited by the *Effort (E)* required to move attention (e.g., a pilot will be less likely to scan an instrument located at an overhead panel, head down, or to the side where head rotation is required, than to an instrument located directly ahead on a head-up display, HUD) (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). Visual perception in MIDAS depends on the amount of time the observer dwells on an object and the perceptibility of the observed object. The SEEV model has been integrated into MIDAS (Gore, Hooey, Wickens, & Scott-Nash, 2009) and drives the operators' visual attention.

The perception model computes the perceptibility of each object that falls into the operator's field of view based on properties of the observed object, the visual angle of the object and environmental factors. In the current implementation of MIDAS, perception is a three-stage, time-based perception model (undetected, detected, comprehended) for objects inside the workstation (e.g., an aircraft cockpit) and a four-stage, time-based perception model (undetected, detected, recognized, identified) for objects outside the workstation (e.g., taxiway signs on an airport surface) (Arditi & Azueta, 1992; Harber & Hershenson, 1980; Lubin & Bergen, 1992; Robinson, 1979; Robinson, Koth, & Ringenbach, 1976). Information then passes into a three-stage memory store (Baddeley &

Hitch, 1974; Ericsson & Kintsch, 1995) that degrades according to empirically driven memory decay rates (Alexander & Wickens, 2005; Card, Moran, & Newell, 1983; Endsley, 1995; Gugerty, 1998).

The cognitive models interact with a series of validated anthropometric models (Badler, Phillips, & Webber, 1993) that call a number of validated motor movement models (Fitts' Law, (Fitts, 1954; Shannon & Weaver, 1949; Welford, 1976).

The **MIDAS output model** (Figure 2, right column) generates a runtime display of the task network, timeline, fit, reach, and visibility envelopes, a dynamic animation of the operator carrying out his/her tasks within the environment, and mission performance measures such as workload and visual fixations.

Tasks are triggered by information that flows from the environment, through a perception model, to a task network representation of the procedures that then feeds back into the environment. Tasks are characterized by several defining parameters that include conditions under which the task can be performed (e.g., beginning, ending, and wait-for clauses), its relative priority with respect to other tasks, an estimate of its duration for scheduling, its interruption specifications, and the resource required to perform the task defined according to the Modified Task Analysis and WorkLoad (TAWL) (McCracken & Aldrich, 1984; Mitchell, 2000).

II. HUMAN PERFORMANCE MODEL OF CURRENT-DAY RNAV OPERATIONS

Developing a model of RNAV Approach and Land Operations

The objective was to develop a high-fidelity model of two-crew (pilot flying, PF, and pilot-not-flying, PNF) commercial transport operations, with ATC tasks and procedures modeled at a lower level of fidelity, but at a level sufficient to represent the interactions between pilots and ATC. The model was based on a scenario in which pilots flew a RNAV approach into Dallas Fort Worth (DFW) with current-day Boeing-777 equipage (see Figure 4). The scenario began with the aircraft at an altitude of 10,000' and 30nm from the runway threshold. The cloud ceiling was 800', with a decision altitude of 650' at which point the modeled pilots disconnected the autopilot and manually hand-flew the aircraft to touchdown.

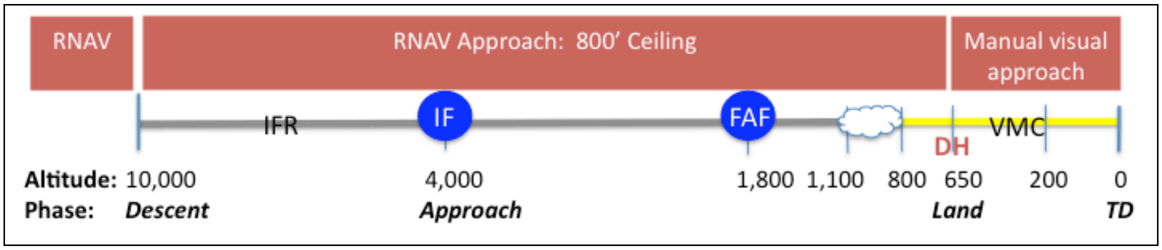


Figure 4. Baseline RNAV Model of Approach and Land.

Notes: DH = Decision Height; FAF = Final Approach Fix; IF = Initial Fix; IFR = Instrument Flight Rules; RNAV = Area Navigation; TD = touchdown; VMC = Visual Meteorological Conditions.

The RNAV model incorporated the series of pilot tasks according to current roles and responsibilities for flying an RNAV approach. The model was based on cognitive task analyses of flight tasks (Gore, Hooey, Wickens, Sebok, et al., 2009; Keller, Leiden, & Small, 2003), and cognitive walkthroughs with a commercial pilot and air traffic controller. This process generated a comprehensive set of tasks in each of the following major phases of flight: Descent, Approach, and Land. During descent (from 10,000' to 4,000'), the PF controls the aircraft autopilot using the MCP and the PNF is primarily responsible for radio calls with approach control, checklists, and crosschecking PF command inputs. During the approach phase (from 4,000' to 650'), the crew slows the aircraft and configures it for landing by progressively lowering flaps and then the landing gear prior to the final approach fix (FAF). At the FAF, the PNF radios Tower Control, as directed, to obtain landing clearance. In the Land Phase (650' to touchdown (TD)) the crew prepares to land the aircraft. After obtaining a visual identification of the runway, the PF disconnects the autopilot and flies the aircraft to the runway, flares to bring the main landing gear to the pavement, and then flies the nose to the runway. The scenario ends when the main gear contacts the runway.

The environment triggers the procedures within the operator model. These procedures are the set of tasks that the operator must follow to complete their goal-related performance (see Figure 5).

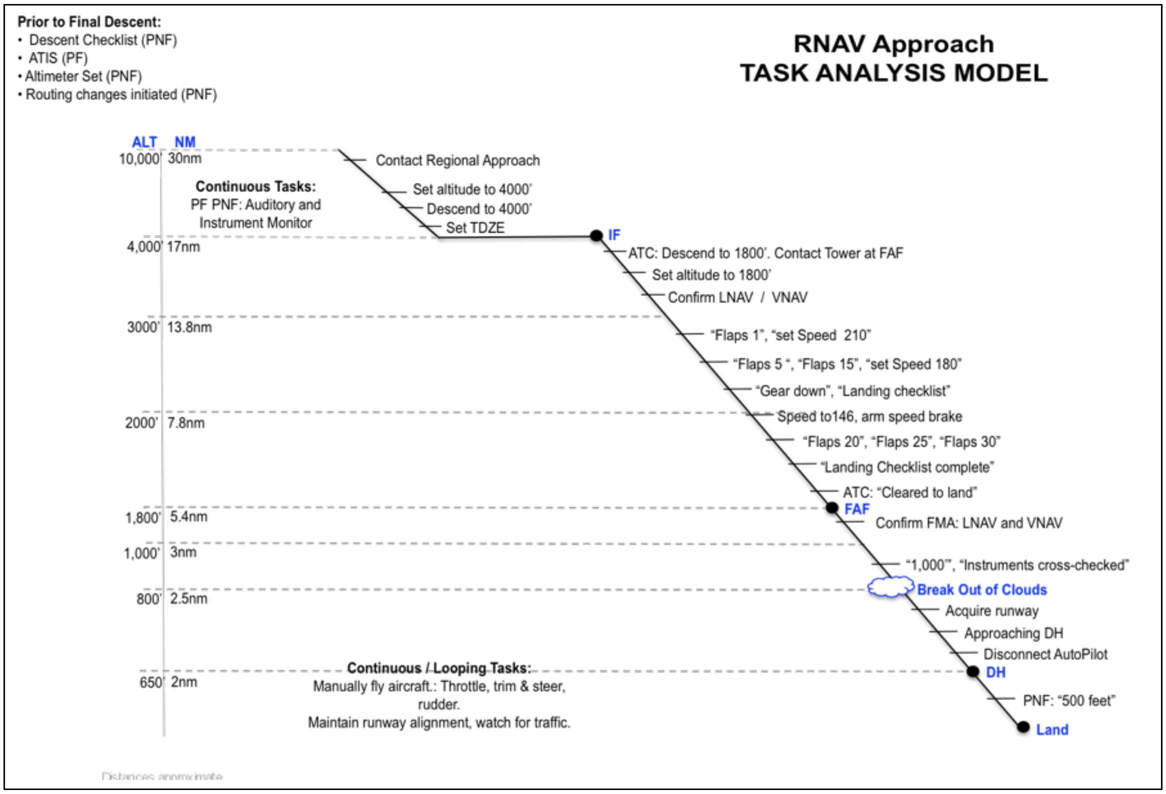


Figure 5. High-level Task Analysis of the RNAV Model.

Notes: ALT = Altitude; ATC = Air Traffic Control; DH = Decision Height; FAF = Final Approach Fix; IF = Initial Fix; LNAV = Lateral Navigation; NM= Nautical Miles; PF = Pilot Flying; PNF = Pilot Not Flying; RNAV = Area Navigation; TDZE = touchdown zone elevation; VNAV = Vertical Navigation

The task model is comprised of major procedures that are then broken down into a set of task primitives at a fine-grained level of fidelity. For example, the task of pressing a button on the MCP is translated into the following sequence of behavioral primitives: *reach, push and release, return arm*. These were then translated into MIDAS' MicroSaint Sharp task network structure. The model was comprised of over 970 tasks that included the environment parameters as well as the flight crew and ATC tasks. An example of the task network implementation of the pilots carrying out a sequence of tasks to set the flaps is presented in Figure 6.

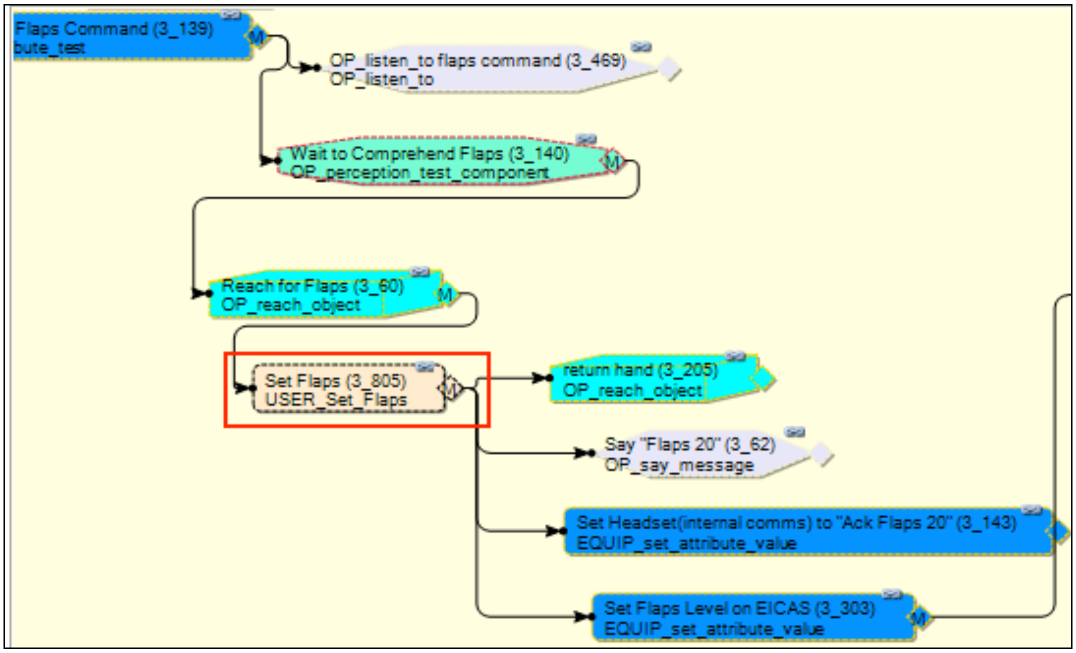


Figure 6. Task Network Model Implementation of a "Set Flaps" Sequence.

The model was then verified by an independent pilot, not involved in the initial model development process, to determine whether the model faithfully represented the pilots' tasks. To verify the model was implemented error-free, a new task analysis was reverse-engineered by analyzing the model output of task begin and end times and compared to the original task analysis. This reverse-engineered task analysis was presented to the subject matter expert for analysis. The reverse engineering process culminated in a list of the tasks, their order, and the associated task time that the task would occur in the approach and land timeline. The focus group pilots assessed the tasks for completeness, sequence, and timing. The model's task network was refined when discrepancies existed. The reverse engineer process served to illustrate the tasks that the model used to generate its output and to develop specifications for model redesign.

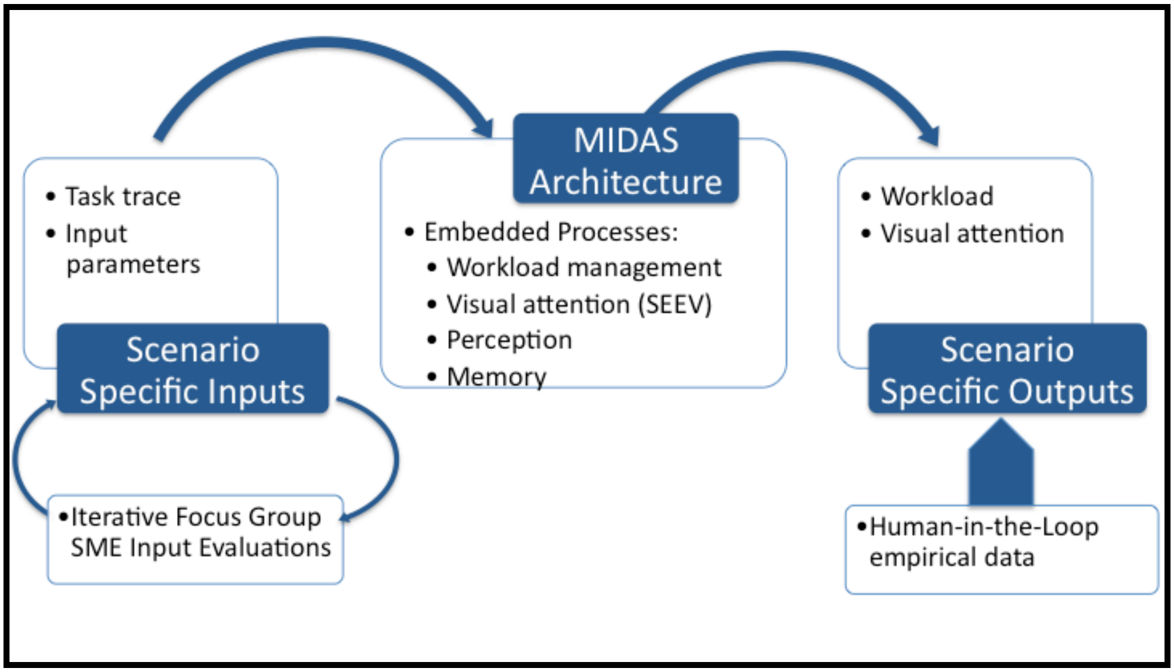
Validating the RNAV Model of Approach and Land Operations

Model validity can be considered from many different angles as revealed by (Gluck & Pew, 2005), Foyle and Hooey (2008), and (Gore, 2002)including evaluation of model inputs and model outputs. Model inputs refer to the task trace and model parameters such as task times and task loads, whereas model outputs refer to operator performance measures, such as task timing, visual fixations, workload, or situation awareness. Gore (2010b) outlines that HPMs of complex operations, should be evaluated using multiple measures and that these measures should be taken from different levels of fidelity. Some model verification and validation efforts rely solely on input validation, while others focus on validating solely the output, termed results validation (Campbell & Bolton, 2005). Relying only on output validation, or results validation, as the sole measure as is frequently the case is insufficient because there is no guarantee that the model is representative of the operators tasks or cognitive processes (Hooey & Foyle 2008).

Indeed, it is possible that one could make parameter manipulations until the model output fits the data, while misrepresenting the sequence or order of tasks, the workload associated with carrying out the individual tasks, or the way the operator processes the information from the environment. In this case, the model does not validly represent the pilot's tasks and may lead to invalid conclusions when the model is extended with new scenarios, tasks, or environments.

The MIDAS approach for model validation is presented in Figure 7. This methodical approach is multi-dimensional using multiple variables at varying levels of resolution. The validation process involves three validation components: validation of the inputs, validation of the process models in the architecture, and validation of the model outputs. The input and output components are scenario-specific while the architecture process models are general and not specific to the domain application model (i.e. perception is perception no matter if the task is driving or flying). Note that validation of the process models themselves, including workload management, perception, and visual attention are important aspects of the validation progression. These process models within MIDAS have been previously validated (as discussed above) and are held constant across domain applications, so will not be discussed further here.

This application model of flight deck operations was validated by a thorough evaluation of ***model inputs***, including the task trace and model input parameters required for the workload and visual attention models and ***model outputs*** including workload and visual fixation (percent dwell time; PDT). The process was iterative, in that the model was refined based on the input validation process and the output was validated by comparing the refined model to empirical human-in-the-loop data.



MIDAS Architecture

- Embedded Processes:
  - Workload management
  - Visual attention (SEEV)
  - Perception
  - Memory

Scenario Specific Inputs

- Task trace
- Input parameters

- Iterative Focus Group SME Input Evaluations

Scenario Specific Outputs

- Workload
- Visual attention

- Human-in-the-Loop empirical data

Figure 7. Comprehensive Model Validation Approach.

Input Validation

Our validation of the model inputs included two aspects. First, a formal validation of the task trace was conducted to determine the extent to which the modeled tasks represent the pilots' actual tasks. Second, a formal analysis was conducted to determine the validity of the model input parameters of workload assigned to the basic task primitives. MIDAS uses behavioral primitives that contain workload estimates based on the Task Analysis Workload (TAWL) index (McCracken & Aldrich, 1984). These values are based on inputs from military rotorcraft pilots, and have not previously been validated by commercial pilots, for the task of conducting approach and landing tasks in fixed-wing aircraft. Other input parameters, including information importance, relevance and criticality for each phases of flight, were also collected but are not reported here.

Method

Two full-day focus group sessions were conducted to evaluate the validity of the model inputs. Each session was comprised of four pilots. The eight pilots (six Captains and 2 First Officers) were current commercial pilots of glass-cockpit aircraft (M=1,317 flight hours; range 250 to 3,000 flight hours), with RNAV-approach experience. After completing the demographics form, the focus group participants were introduced to human performance modeling and the high-level goals and objectives of the focus group session. Using a scenario-based format, the focus group pilots conducted a cognitive walkthrough of a typical approach and land scenario, starting at 10,000' and continuing to touchdown. Pilots were asked to consider the tasks required of the PF and the PNF, and the nature of communications between ATC and the pilots.

The PF, PNF and ATC tasks as modeled in MIDAS were presented on a worksheet to the participants and each pilot was asked to independently review the tasks. They were asked to identify any tasks that were assigned to the incorrect operator, occurred in the incorrect sequence, or at an incorrect altitude or navigation marker, and identify any tasks that were missing. Upon completion of the worksheet, the pilots discussed their evaluations in a semi-structured round-table format and identified the source of any discrepancies among pilots. Each session concluded with the pilots reaching consensus on a set of tasks for PF, PNF, and ATC.

Next, pilots were trained to estimate task workload along five dimensions (Visual, Auditory, Cognitive, Speech and Motor) using the 7-point TAWL scale with behavioral anchors (Hamilton, Bierbaum, & Fullford, 1990; McCracken & Aldrich, 1984). The focus group participants were asked to first identify the workload dimensions that were applicable for the given task, and then estimate the workload for each relevant dimension using the 7-point scale. Two categories of behavioral primitives were evaluated: 1) Basic behavioral primitives previously validated by McCracken & Aldrich, such as reach, button press, and manipulate toggle switches; and, 2) RNAV model-specific behavioral primitives, which had not been previously validated, such as "acquire lead aircraft", "track lead aircraft", "visually acquire runway", "monitor runway alignment", and "verify FMA readout". The basic behavioral primitives served as a baseline upon which to evaluate whether the focus group pilots' workload estimates were comparable to the previously validated TAWL scale.

Results

Task Trace Validation. The task trace worksheets were reviewed among the group of four pilots in each focus group session. Each pilot described his recommended changes and sources of the discrepancies among the pilots' worksheets were discussed. Differences among tasks were attributed to differences due to aircraft type, airline, pilot technique, and airport/airspace procedures. Out of 74 tasks in the MIDAS RNAV application presented, the focus group pilots identified 12 tasks that should be removed, reordered, or added. In addition, pilot-ATC phraseology was refined to better reflect actual operations. Incorrectly representing the communication length or the information contained within the communication results in misestimates of workload and task time to reach comprehension. The MIDAS input model was modified to reflect these changes. The refined model more accurately represented the information being communicated as well as the task sequence.

Input Parameter Validation. For each task, the mean estimated workload value for each workload dimension was compared to the MIDAS input parameters, with the constraint that at least six of the eight focus group participants determined that the dimension was relevant for the task. One sample t-tests were conducted to compare the mean focus group rating to the MIDAS value. Significant results indicated that the pilots' estimated workload values were significantly different than the MIDAS values. Thirty-nine tasks were rated on the visual, auditory, cognitive, and motor dimensions (as relevant for the task) resulting in 75 ratings.

Three of the behavioral primitives that the subjects were asked to rate were taken directly from the previously validated TAWL index. They were *push-and-release, reach object, and say message*. For all three primitives, the focus-group mean ratings did not differ significantly from the TAWL ratings (Push-and-release - $t_{visual}(7)=4.2$, $p>.05$, t_{motor} $(7)=12.19$, $p>.05$; reach – $t_{visual}(7)=.306$, $p>.05$, $t_{cognitive}$ $(7)= 1.17$, $p>.05$, t_{motor} $(6)=1.37$, $p>.05$; say - $t_{cognitive}$ $(7)=.877$, $p>.05$). This is evidence that these TAWL ratings, though validated on an army rotorcraft task, do extend to fixed-wing commercial aircraft. These results also suggest that the focus-group pilots were trained sufficiently on the TAWL scale to produce answers in accordance with the TAWL, thus providing confidence in the pilots' ratings for the non-TAWL primitives, as discussed next.

Our initial model mapped four of the pilots' high-level tasks (*set speed, set flaps, set gear, and tune radio frequency*) to the *push-and-release* task primitive. However, the focus-group ratings of these input parameters revealed that each possessed unique workload properties that differed significantly from the push-and-release TAWL values (sest speed - $t_{visual}(7)=3.5$, $p<.05$; $t_{cognitive}(6)=1.7$, $p<.05$; Set flaps - $t_{motor}(7)=4.5$, $p<.05$; set gear - $t_{motor}(7)=4.9$, $p<.05$; tune radio frequency - $t_{visual}(7)=4.2$, $p<.05$; $t_{motor}(7)=12.1$, $p<.05$;). For example, *set-speed* required more cognitive workload and more motor workload than *push & release*. Unique task primitives were developed for each of these four pilot tasks using the mean focus-group ratings.

The baseline RNAV model required three new primitives, not contained in the TAWL, which were specific to approach-and land operations in commercial fixed-wing aircraft.

These were: *visually acquire runway; manipulate yoke; manipulate pedals*. The mean focus-group ratings were used for the relevant workload parameters.

Table 1. RNAV-Specific Task Mean Ratings.

Task	Visual	Auditory	Cognitive-Spatial	Cognitive-Verbal	Motor	
					Fine	Gross
Acquire runway	5		3.7			
Manipulate yoke	1.2		1		1.3	1.3
Manipulate pedals			1			3.2

Output Validation

In the output validation phase, the model outputs of workload and dwell percentage were compared to empirical data from the existing literature. This phase was completed only after all of the inputs into the HPM were modified based on the task trace and parameter input analyses described previously.

Method

The baseline RNAV model's predicted workload and PDT data for the descent (10,000'-1,800'), approach (1,800'-650'), and land (650'-TD) phases of flight were compared to empirical data from independent HITL simulations available in the literature. Statistical correlation tests were conducted to evaluate the goodness-of-fit between the model and the HITL data. For all analyses, only the PF data are shown.

A survey of the literature was conducted to identify relevant HITL data sources that provided workload and/or visual fixation data from commercial pilots flying approach-and-land scenarios in a glass cockpit in either an actual flight test or a high-fidelity flight simulator. One HITL study was identified as a suitable comparison for the workload data (Hooey & Foyle, 2008). This medium-fidelity HITL simulation was previously conducted for a different model validation effort, and as such was unique in that it provided workload estimates from three commercial pilots using the TAWL scale, for the three phases of flight modeled in the current baseline RNAV model. Three additional HITL studies (Anders, 2001; Hüttig, Anders, & Tautz, 1999; Mumaw, Sarter, & Wickens, 2001) were identified as suitable comparisons for the visual fixation data. Each study included commercial pilots, flying ILS approach-and-land scenarios.

Results

Workload. Figure 8 presents the overall workload as predicted by the MIDAS RNAV model (solid line) and the mean of the three pilots estimates (dashed line) in the Hooey & Foyle, (2008) HITL simulation for each of three phases of flight (descent, approach, and land). An average of the individual channel workload values (visual, auditory, cognitive and motor) was taken for the HPM at each timestamp. The timestamp average was then averaged across the phase of flight. As can be seen, the HPM and HITL data exhibit the same trend (r^2=.54). The model data tended to over-predict workload during the landing phase. It should be noted that the HPM simulation was a medium-fidelity simulation, with only three participants, and lacked the high fidelity representations of the instrumentation and controls, which may have actually lead to an under-estimation of workload by the actual pilots in the HITL simulation.

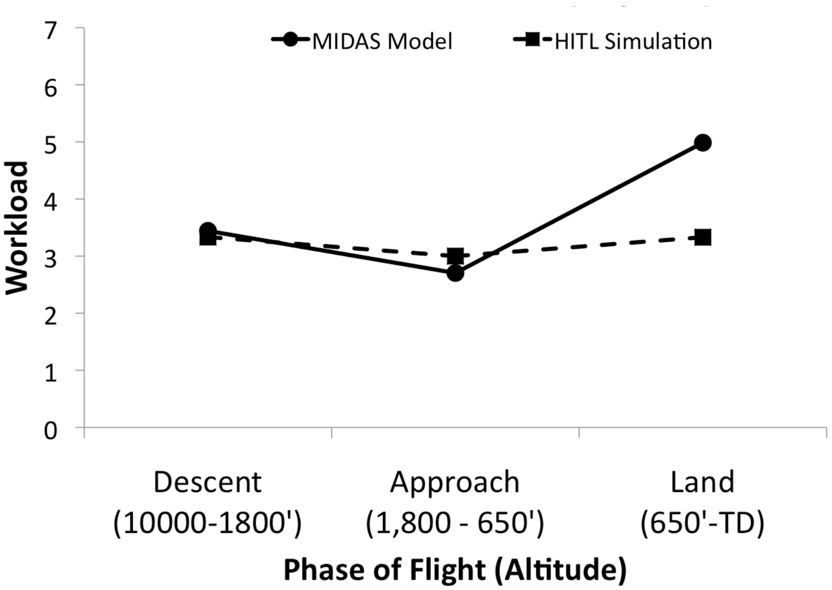


Figure 8. Pilot Flying's Overall Workload from MIDAS model and Human-in-the-loop Simulation.

In order to evaluate the operation of the model more fully, the Hooey & Foyle (2008) workload output for the Pilot Flying (PF)[5] was compared to the RNAV model's workload predictions of visual, auditory, cognitive, and motor channels.

Visual Workload.

Visual workload refers to the visual demands associated with a set of tasks. Figure 9 illustrates that the HITL and model visual workload output show a similar pattern of results (r^2=0.55) with visual workload being less during the approach phase than either descent or land. The model predicts a larger increase in visual workload during the land phase of flight than shown in the HITL data.

---

[5] In Hooey & Foyle (2008), the PNF was a confederate pilot and visual fixation data were not collected

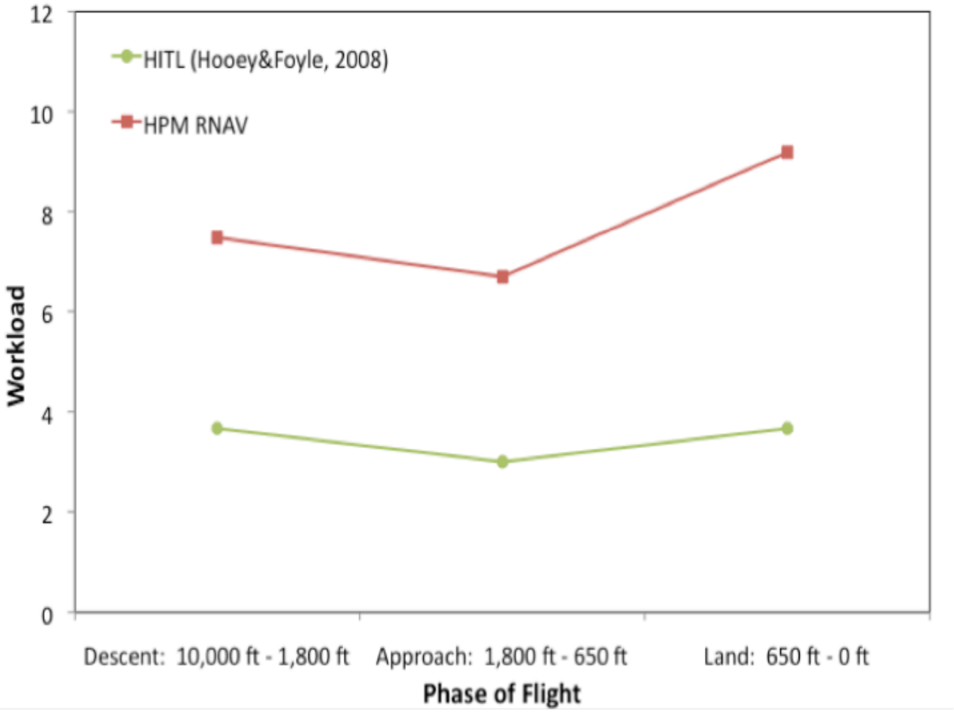


Figure 9. Pilot Flying Visual Workload. Data Shown are HITL Data (from Hooey & Foyle, 2008) and RNAV HPM Predictions.

Auditory Workload. Auditory workload refers to the auditory demands associated with a set of tasks. Figure 10 illustrates that the model predicts a pattern of results similar to the HITL data, r^2=0.76, with auditory workload declining as the aircraft moves from decent to approach to land. This means that the auditory workload was impacted by the phases of flight in a similar fashion in both the model and the HITL data.

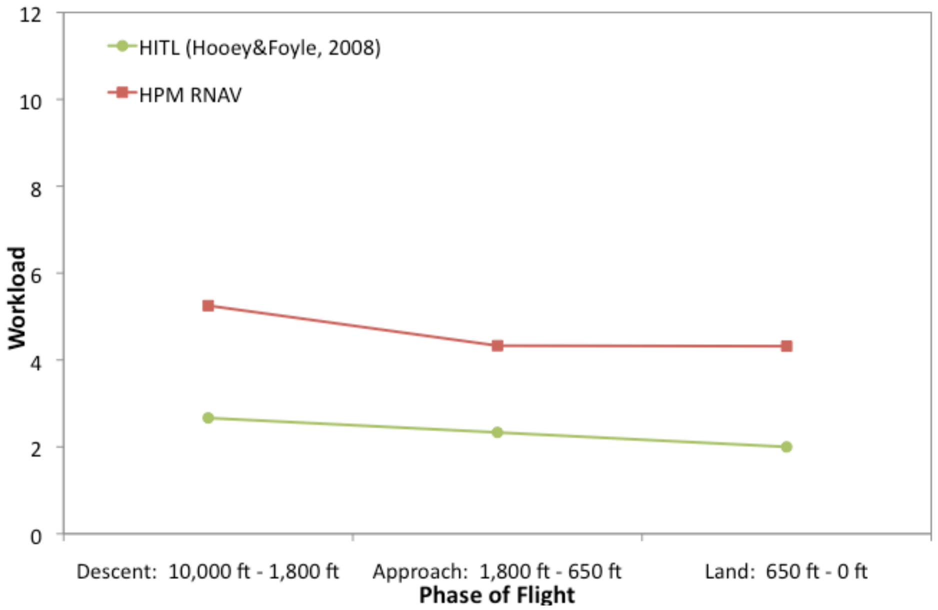


Figure 10. Pilot Flying Auditory Workload. Data Shown are HITL Data (from Hooey & Foyle, 2008) and RNAV HPM Predictions.

Cognitive Workload. Cognitive load refers to the cognitive demand associated with the tasks that comprise the phase of flight. In the HITL study, pilots estimated 'cognitive' workload, and as presented in Figure 11, the data show a decline in cognitive workload when moving from the descent to the approach phase of flight followed by an increase during the land phase of flight. In the model, however, cognitive workload is predicted along two dimension: Cognitive-verbal and cognitive-spatial. Figure 11 shows both the verbal and spatial dimensions separately and as an average cognitive load. The cognitive-verbal model data shows a decline in cognitive workload when moving from the descent to the approach phase of flight followed by an increase in cognitive workload in the land phase of flight. The trend associated with this pattern is correlated to the HITL data (r^2=0. 61). The cognitive spatial workload and the average of the cognitive-spatial and cognitive-verbal model data presented in Figure 11 shows a similar decline in cognitive workload when moving from the descent to the approach phase of flight followed by a very large increase in cognitive workload during the land phase of flight (r^2=0.18, and r^2=0.24 respectively). That the HITL pilots estimate of cognitive workload is most highly correlated with the model's predicted cognitive-verbal output, suggests that humans tend to overweight the cognitive-verbal elements as compared to the cognitive-spatial elements in their estimation process.

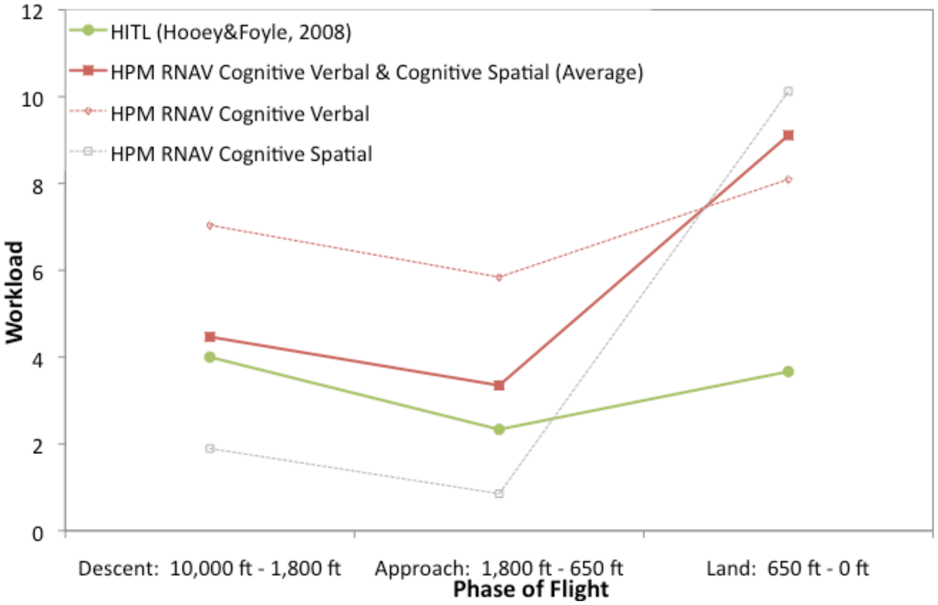


Figure 11. Pilot Flying Cognitive Workload. Data Shown are HITL Data (from Hooey & Foyle, 2008) and RNAV HPM Predictions.

The Motor Channel. Motor workload refers to the motor demands associated with a set of tasks that comprise the phase of flight under examination. Figure 12 shows the HITL estimates of workload and the HPM predictions, averaged over gross motor, fine motor, and voice. Both the HITL data and the model predictions are characterized by a decline in motor workload when moving from the descent to the approach phase of flight followed by an increase in motor workload in the land phase of flight. These trends are highly correlated (r^2=0. 93).

Figure 12. Pilot Flying Motor Workload. Data Shown are HITL Data (from Hooey & Foyle, 2008) and RNAV HPM Predictions.

Visual Fixations (Percent Dwell Time). Figure 13 presents the model fit of PDT on three areas of interest: Primary Flight Display (PFD), Navigation Nav) Display, and Out-the-Window (OTW). The model PDT output was compared to the three separate HITL data sets. There was a strong positive correlation $r^2 = .99$) between the RNAV model PDT output and the average of the three HITL studies (Anders, 2001; Hüttig, et al., 1999; Mumaw, et al., 2001). This strong positive correlation is evidence that both the model inputs and the SEEV process model, which guides visual attention, are valid.

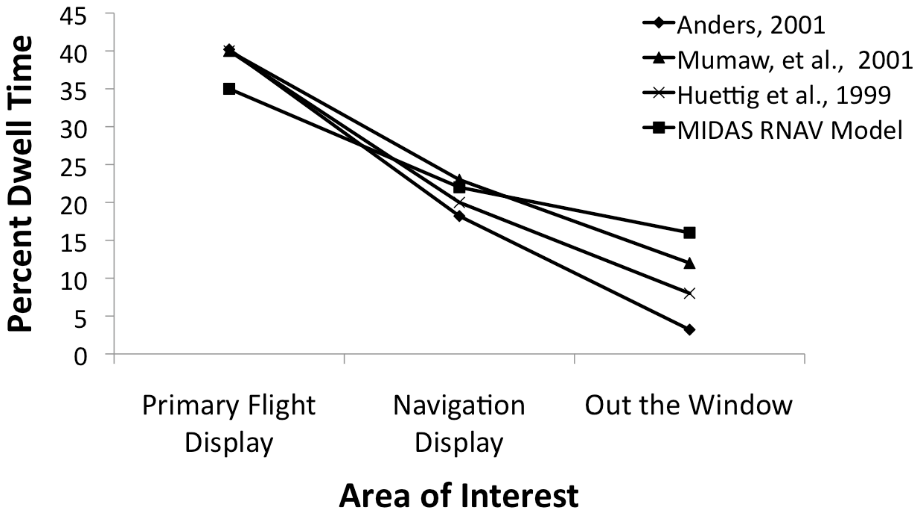


Figure 13. Output Validation: PF Visual Fixations from MIDAS RNAV Model Compared to Three HITL Simulations.

DISCUSSION AND SUMMARY

An application model of commercial airline pilots conducting approach-and-land procedures was created using the MIDAS software following a methodical development and validation approach. The premise that guided the current work was that model validity is a process, not solely a single value at the conclusion of a model development effort. Valid inputs lead to valid outputs. It is therefore necessary to follow an iterative input validation process as well as an iterative output validation process. Conducting only one of these validation processes may lead to invalid models. This is especially true as the complexity of the operational environment and tasks increase.

The pilot focus groups were instrumental in defining valid model inputs. The scenario-based cognitive walkthrough approach captured the context of operations well and enabled the pilots to easily identify tasks that depend on specific phases of flight, and augment the environmental considerations that are used to drive the model's performance.

The workload associated with the behavioral primitives was evaluated with some degree of success. This effort illustrated that the base set of MIDAS workload primitives, derived directly from the TAWL, were valid as evaluated by the focus-group pilots. Context-specific workload primitives were modified based on pilot input.

The model output correlated strongly with multiple independent human-in-the-loop simulation studies. These output validation results provide further evidence that the model inputs and the workload and SEEV process model are valid.

In summary, the methodical and comprehensive model validation effort presented in this paper illustrates a candidate process for developing and validating HPMs. The credibility of the model is improved when a rigorous validation effort is followed that extends beyond output validation. This valid current-day RNAV model will next be extended to evaluate the impact of potential NextGen CSPO concepts.

III. EXTENDING THE VALIDATED RNAV MODEL TO CLOSELY SPACED PARALLEL OPERATIONS

The next phase is to extend the validated RNAV model to NextGen CSPO scenarios to evaluate the impact that the procedural changes will have on pilot workload and visual attention.

A HUMAN PERFORMANCE MODEL OF CLOSEL SPACED PARALLEL OPERATIONS

The validated RNAV scenario was modified to reflect the anticipated changes to the rules and procedures associated with the VCSPA concept (see VCSPA section on page 2). Two operational implementations (VCSPA 800' and VCSPA 200') of the VCSPA concept were modeled. The VCSPA 800' condition assumed the same operational environment as the RNAV model with a cloud ceiling of 800' and DH of 650' and manual land procedures. The VCSPA 200' condition, assumed an operational environment more consistent with NextGen goals, with a lower landing minima, specifically, a cloud ceiling of 200' and a DH of 100' (see Figure 14).

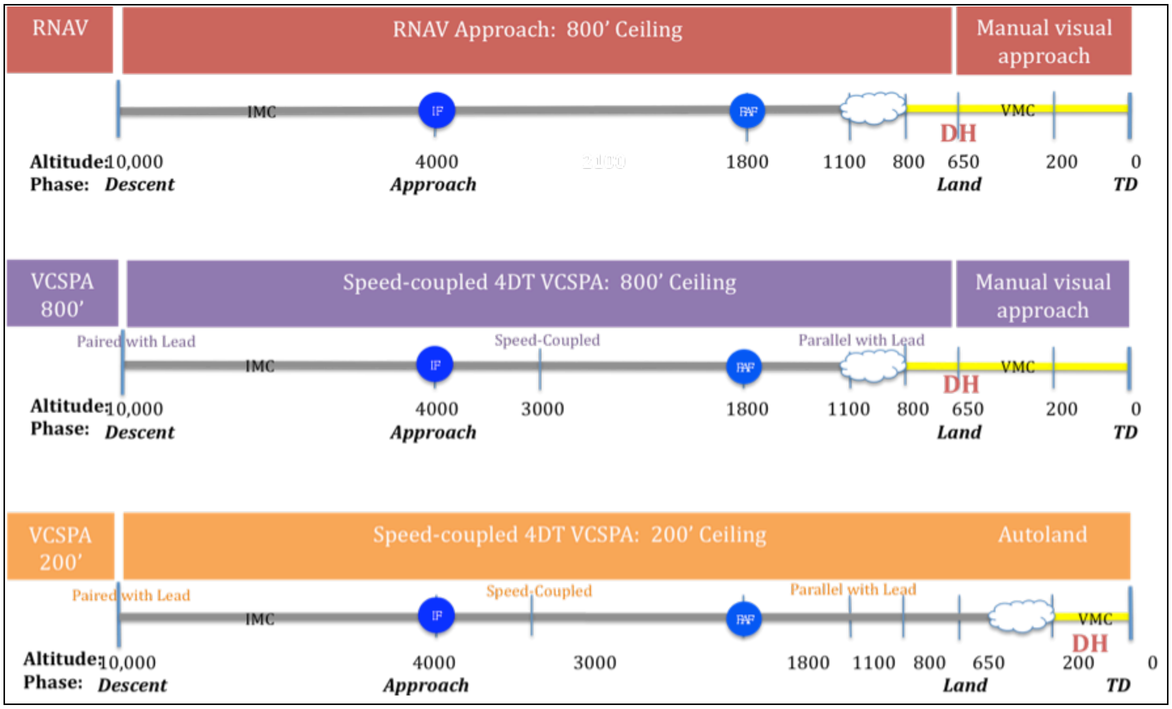


Figure 14. RNAV, VCSPA 800' and VCSPA 200 scenario specifications.

Notes: 4DT = Four-dimensional Trajectory; DH = Decision Height; FAF = Final Approach Fix; IF = Initial Fix; IFR = Instrument Flight Rules; RNAV = Area Navigation; TD = touchdown; VCSPA = Very Closely Spaced Parallel Approach; VMC = Visual Meteorological Conditions.

Both VCSPA implementations assumed the ownship (modeled aircraft) was paired with a lead aircraft that was aligned with a parallel runway prior to scenario start, and pilots engaged speed-coupling automation at 3,000’. At scenario start, the ownship was offset from the runway, and turned to align with the runway at approximately 1,100’. The NextGen cockpit was equipped with wake displays consistent with Verma et al. (2008; see CSPO section on page 2) that portrayed wake information on both the PFD and Nav Display.

The basic flight tasks of the RNAV model, such as checklist, flaps, gear, and communication procedures, were held constant across all three models. The VCSPA procedures were branched off of the RNAV procedures so that any difference in model output can be assumed to be due to the different procedures in the RNAV and VCSPA approaches.

For comparison, the model assumptions for the three models are presented below in Table 2.

Table 2. Model Assumptions.

Operational Parameter	**Model**		
	RNAV – 800’	**VCSPA – 800’**	**VCSPA – 200’**
Cloud Ceiling	800’	800’	200’
DH	650’	650’	100’
Paired with traffic	No	Yes	Yes
Wake display on PFD and ND	No	Yes	Yes
Autopilot	Discontinue @ 650’	Discontinue @ 650	Autoland
4D FMS for speed management	No	Yes	Yes

VALIDATING THE CLOSELY SPACED PARALLEL OPERATION MODELS

Method

The VCSPA scenarios were developed by modifying the previously validated RNAV model. As such, in order to ensure the validity of the VCSPA models, the specific VCSPA task changes and input parameters were validated using the same pilot focus group sessions described previously. In the focus group sessions, after the pilots completed the task trace and input parameter worksheets for the RNAV model, the VCSPA concept was introduced. The pilots were briefed on the goals of NextGen, expected changes to flight deck equipage, and pilot procedures. Examples of the wake displays on both the PFD and Nav Dispaly and the visual and auditory wake warnings and alerts were presented. A video of two pilots completing VCSPA procedures from Verma et al. (2008) was also presented. Next, the VCSPA 800’ implementation was introduced and pilots were briefed on the operational assumptions. Pilots completed the same task trace worksheet and input parameter rating sheets as conducted for the RNAV model. This was repeated for the VCSPA 200’ implementation.

Results

Task Trace Validation.

The task trace worksheets were reviewed among focus group pilots in each session. Each pilot described his recommended changes and sources of the discrepancies among the pilots' worksheets were discussed. Differences among tasks were attributed to differences due to aircraft type, airline, pilot technique, and airport/airspace procedures. Out of 89 MIDAS tasks presented, the focus group pilots identified 10 tasks that should be removed, reordered, or added. A key finding from the focus group session was concern that the coupling task at 12 nm from the runway threshold was too late. During the coupling procedure, pilots receive a datalink from ATC and then engage the automation using the MCP. Some pilots thought this should occur much earlier, at least 20 nm from the runway threshold, but others warned that the aircraft would not be fully configured at this point. This remains an issue for future research, but the coupling point remained at 13 nm for the current models. In addition, 12 pilot-ATC communication events were refined to better reflect actual operations. The MIDAS input model was modified to reflect these changes. The refined model more accurately represented the information being communicated as well as the task sequence.

Input Parameter Validation.

Three new tasks, specific to VCSPA operations, were created and the focus group pilots were asked to determine which workload parameters were relevant for the task, and then estimate the workload using the 7-point TAWL scale. (Recall, that pilots had previously been trained on the TAWL scale's behavioral anchors). The mean value for each workload dimension was adopted, with the constraint that at least six of the pilots determined the workload dimension was relevant. The three VCSPA-specific tasks are presented in Table 3 along with the cognitive channel statistical test completed on the average pilot rating as compared to the baseline MIDAS primitive value.

Table 3. VCSPA-Specific Task Mean Workload Ratings.

Task	Visual	Auditory	Cognitive-Spatial	Cognitive-Verbal	Motor
Acquire Lead Aircraft	5		5.69		
Read datalink	5			3.00	
Track Lead Aircraft	5		5.68		

MODEL RESULTS

The previously validated RNAV model's predicted workload and PDT data for the descent (10,000'-1,800'), approach (1,800'-650'), and land (650'-TD) phases of flight were compared to the VCSPA 800' and VCSPA 200' scenario. Differences between the baseline model and the VCSPA models can be attributed to different procedures required in each VCSPA implementation.

Workload

Figures 15a through 15d show the model-predicted workload for the PF in the RNAV, VCSPA 800’ and VCSPA 200’ scenarios. As can be seen, for all four workload dimensions, the model predicted negligible differences across the three scenarios for the descent and approach phases of flight. During the land phase of flight, however, the model predicted that the VCSPA 800’ implementation would increase workload across all four workload dimensions. In contrast, there is evidence to suggest that the advanced automation available in the VCSPA 200’ condition (specifically, the autoland capability) will decrease pilot workload in the visual, cognitive, and motor channels.

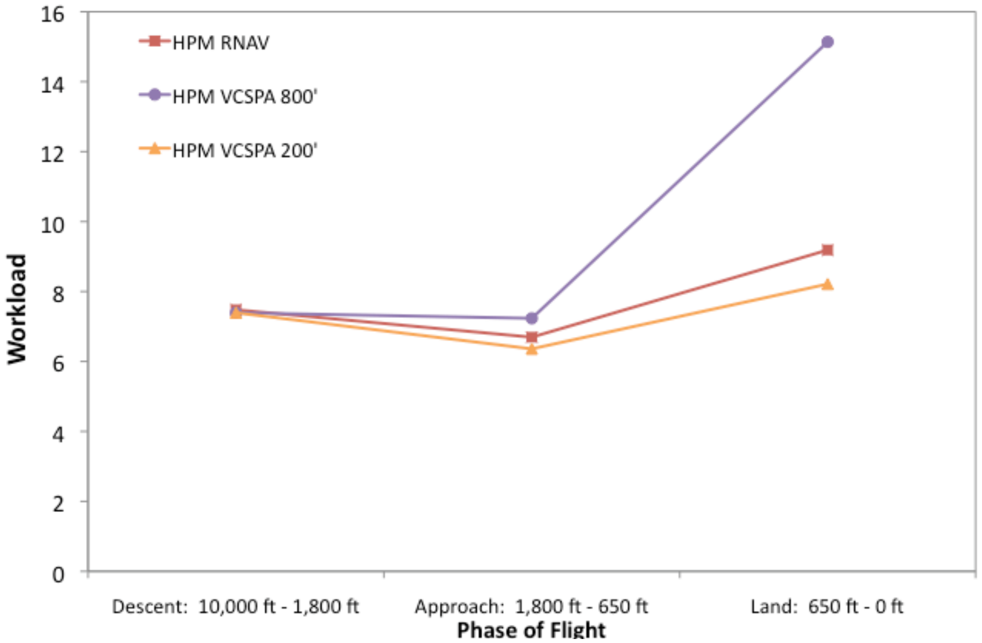


Figure 15a. Model-predicted Visual Workload for the Pilot Flying in the HPM RNAV (Red), VCSPA 800' (Purple), VCSPA 200' (orange) Scenarios.

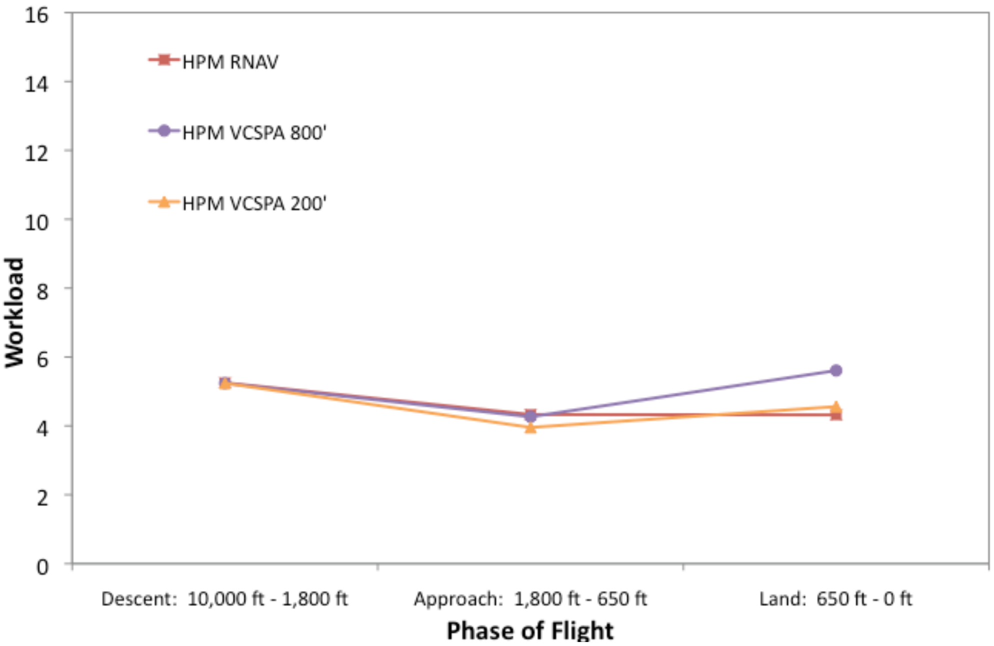


Figure 15b: Model-predicted Auditory Workload for the Pilot Flying in the HPM RNAV (Red), VCSPA 800' (Purple), VCSPA 200' (orange) Scenarios.

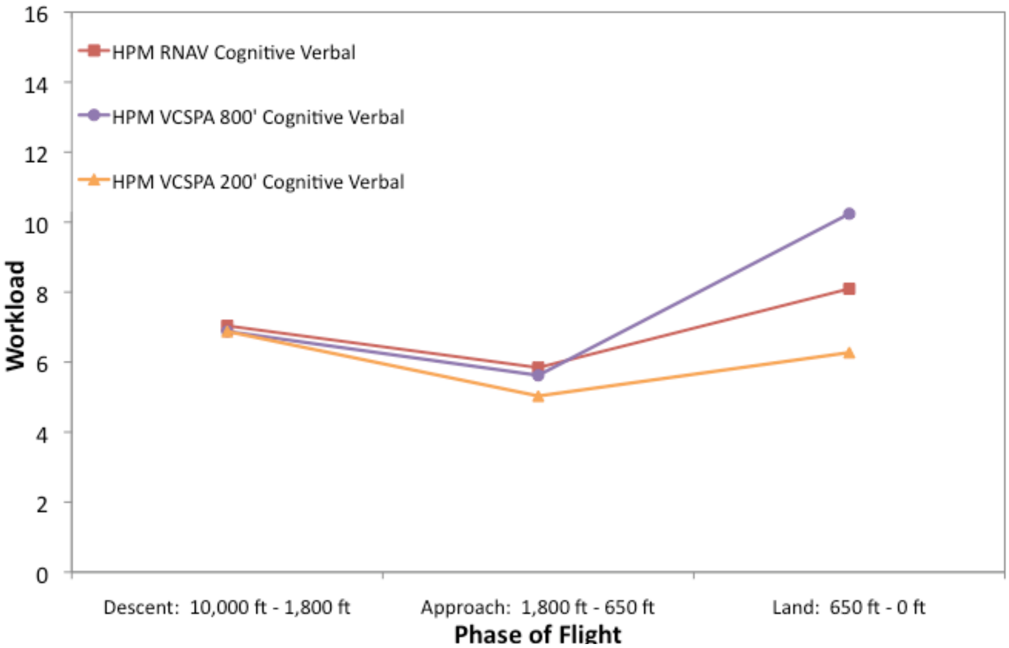


Figure 15c. Model-predicted Cognitive Workload for the Pilot Flying in the HPM RNAV (Red), VCSPA 800' (Purple), VCSPA 200' (orange) Scenarios.

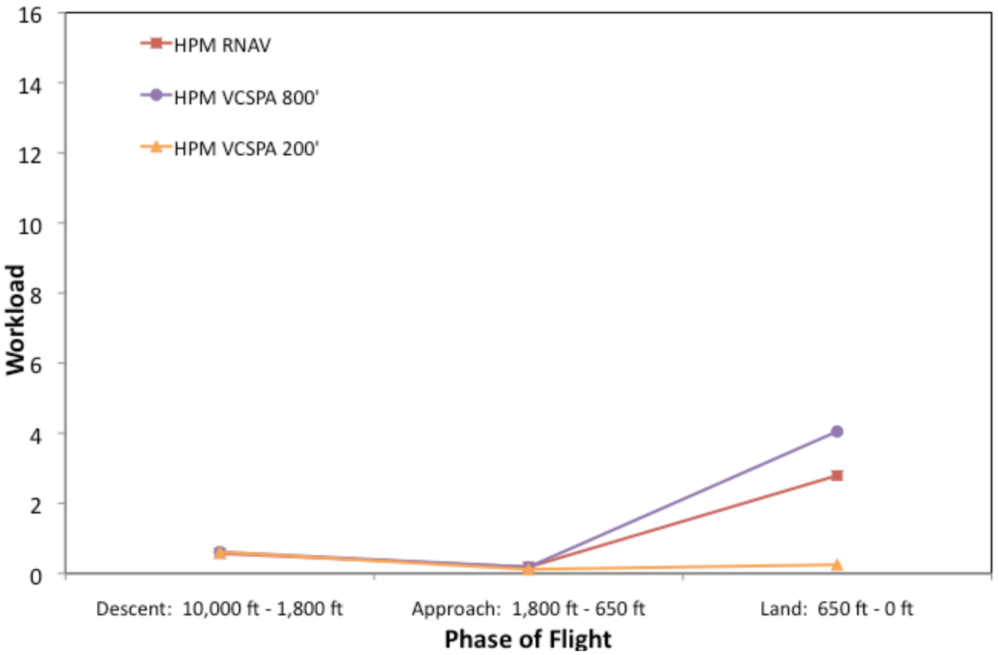


Figure 15d. Model-predicted Motor Workload for the Pilot Flying in the HPM RNAV (Red), VCSPA 800' (Purple), VCSPA 200' (orange) Scenarios.

Visual Attention

Figure 16 presents the PDT to the three main areas of interest (PFD, NavDisplay, and the OTW) as predicted by the model for each of the three scenarios (RNAV, VCSPA 800’ and VCSPA 200’). It is evident by looking at the data presented in Figure 16 that, VCSPA operations are predicted to change the pilots’ visual scan patterns. According to these model outputs, VCSPA will result in an increase in the time spent attending to the Nav Display at a cost of attending to the PFD and OTW.

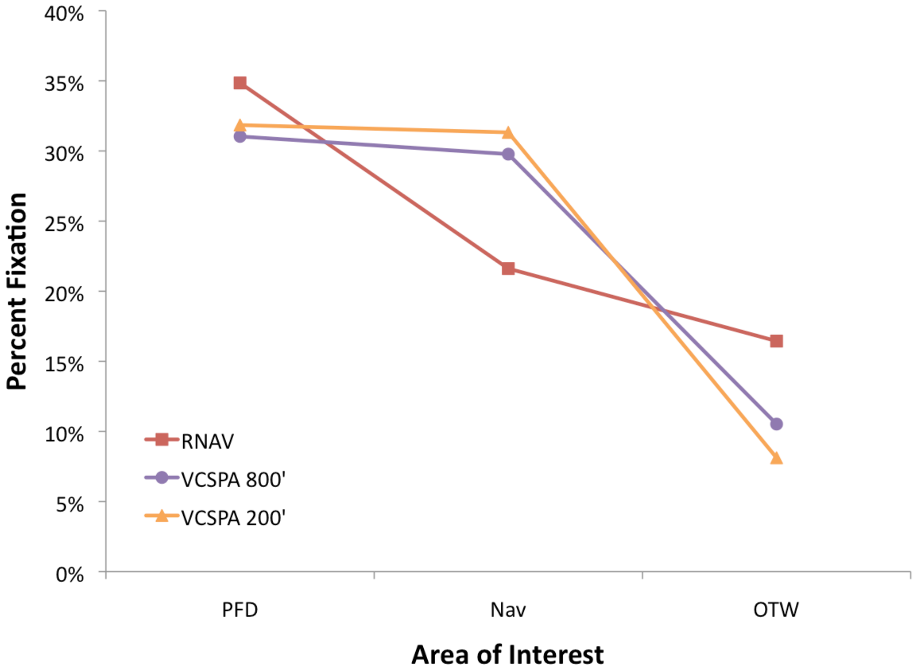


Figure 16. Model-predicted Percent Dwell Time for the Pilot Flying in the HPM RNAV (Red), VCSPA 800' (Purple), VCSPA 200' (orange) Scenarios.

DISCUSSION, IMPLICATIONS AND CAVEATS

The models predict that compared to current-day RNAV approaches, the VCSPA 800' implementation will increase visual, auditory, cognitive, and motor workload on landing. This is not surprising given the high demands for precision hand-flying in the VCSPA 800' condition, and consistent with focus group pilots' subjective impression of the VCSPA 800' implementation. Further, MIDAS predicts that the VCSPA 200' implementation may reduce cognitive and motor workload as compared to the RNAV and the VCSPA 800'. This reduction in workload is attributed to the presence of advanced automation that enables auto-land. Finally, it is apparent that both VCSPA scenarios draw the pilots' attention to the traffic and wake information on the Nav Display at the expense of reduced time on the PFD and OTW.

Implications and Caveats.

1. The workload findings suggest that the VCSPA procedures require advanced automation to support auto-land capability (or at least lower DH) and pilot procedures should prohibit manual approaches in VCSPA conditions. However, it is important to note, that these model results do no account for known, and potentially negative, effects associated with increased automation such as inappropriate trust in automation, complacency and operator-out-of-the-loop phenomenon which may introduce other safety considerations (Dixon & Wickens, 2003; Parasuraman & Riley, 1997). Further work should be conducted

to determine to evaluate how these factors and workload tradeoff. This will be examined further in the following section (off-nominal what-ifs).

2. The visual fixation predictions could be indicative of a vulnerable portion of the approach and land operations in the NextGen. The model predicts that the operator's attention will be spent in the cockpit to a greater extent that outside of the cockpit, which in turn may result in safety implications. For instance, the more time that the pilot spends inside the cockpit, the less likely it is that they will be able to detect an unexpected event in the world – such as aircraft of the runway. This will be examined further in the following section (off-nominal what-ifs).

IV. WHAT-IF SCENARIOS: OFF-NOMINALS

With validated models of RNAV and VCSPA operations reported in the previous section, "what-if" scenarios can now be conducted to explore the impact of these NextGen concepts on pilot's performance during off-nominal events. A comprehensive analysis was conducted to identify appropriate off-nominal events. First, the eight pilots in the focus group session (described earlier in Section II) were asked to identify potential off-nominal events after a scenario-based walk-through of RNAV and VCSPA approaches. Pilots were invited to brainstorm possible off-nominal events and then answered structured probe questions. The pilots generated 21 potential off-nominal events. Second, a search of the ASRS database yielded 199 incident reports, which was narrowed down to 13 potential off-nominal events. Third, findings from previous research (Gore, Hooey, Wickens, Sebok, et al., 2009) was adopted that included a systematic approach to identifying off-nominal events and their contributing factors, and characterizing off-nominal causal factors using a modification of the (Hooey & Foyle, 2003) taxonomy: Environment, System Management, Human and Machine. This yielded 13 potential off-nominal events during approach and land. From these, four off-nominal events were selected. For each off-nominal event, the baseline models presented earlier (RNAV, VCSPA 800' and VCSPA 200') were modified to include the off-nominal event. Each off-nominal model was run in monte carlo mode, for a total of 10 runs. These results are presented next. For all statistical tests, a liberal alpha of .1 was chosen given the exploratory nature of these model evaluations early in the concept development phase.

SCENARIO 1: HIGH WIND / TURBULENCE

The VCSPA 800' and 200' scenarios were modified to represent high-wind, or high-turbulence, conditions. Modifications included increased manual workload associated with the manual flying task, and an increase in the 'expectancy' of flight-relevant areas of interest (AOI) including OTW, PFD, and the wake displays. Also, in NextGen VCSPA conditions, an aircraft may be more likely to deviate from it's flight path, resulting in frequent 'yellow warnings' on the Wake Display (Nav Display and PFD). A pilot might be expected to monitor the paired traffic more closely under these weather conditions. This scenario will demonstrate the effect of this weather condition by comparing the PF's visual attention (PDT) for three major displays (PFD, Nav Display, and OTW) during the land phase (650' to touchdown) in nominal operations and high wind operations in both the VCSPA 800' and 200' conditions. The land phase was chosen for evaluation, because during this phase the paired aircraft are parallel, and closely spaced, thus the effects of high-wind carry the most severe consequences.

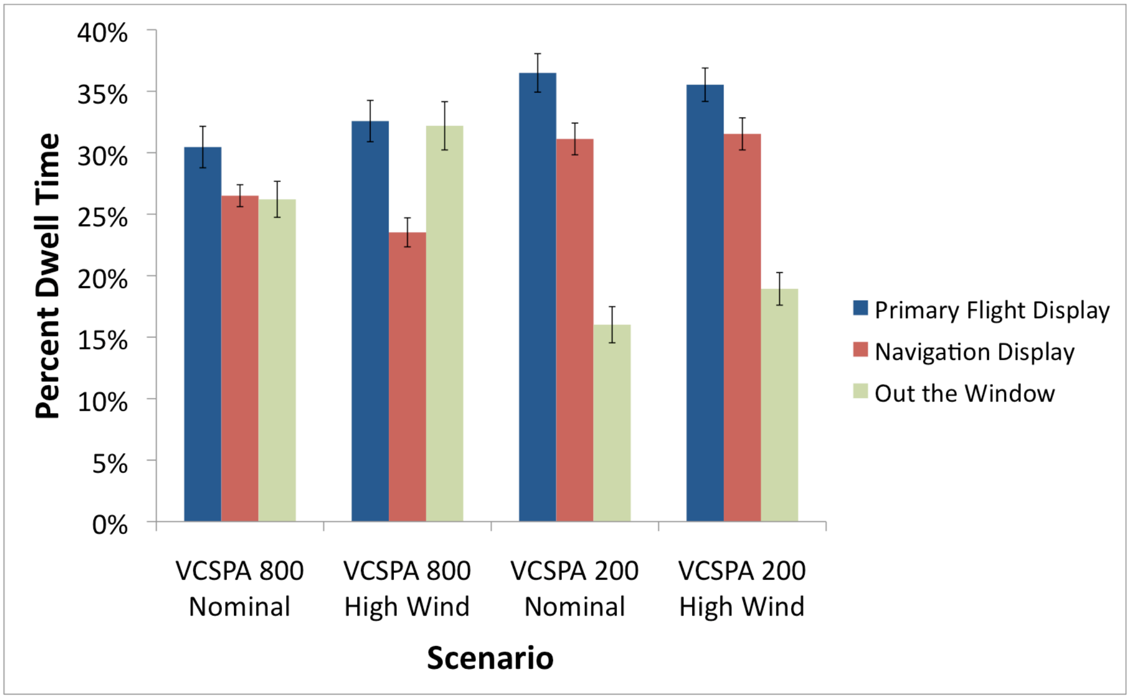


Figure 17. Percent Dwell Time (PDT) to Primary Flight Display, Navigation Display, and Out the Window for the Pilot Flying during the Land Phase of Flight (650' to Touchdown). Graph presents the mean of 10 monte carlo simulation runs with +/- 1 standard error.

Figure 17 presents the PDT for the three main displays within the cockpit, PFD, Nav Dispaly, OTW. When not viewing one of these three displays, pilots were viewing the Upper EICAS, datalink display, FMS or MCP (not shown in Figure 17).

A 2(Scenario) X 2(Wind) X 3(AOI) repeated measures ANOVA was conducted yielding a Scenario X AOI interaction, F(2, 18)=20.52, p<.001. Visual attention to the PFD was higher in the VCSPA 200’ condition (36.01%) than the VCSPA 800’ condition (31.52%), t(9)=2.081, p<.1. Similarly, visual attention to the Nav Display was higher in the VCSPA 200’ condition (31.3%) than the VCSPA 800 condition (25.01%), t(9) = 5.132, p<.01. In contrast, the PDT OTW was higher in the VCSPA 800’ condition (29.2%) than the VCSPA 200’ condition (17.47%), t(9)=6.159, p<.001. This is consistent with expectations because in the VCSPA 200’ condition, the aircraft doesn’t break through the cloud ceiling until 200’ thus the pilots rely more on their instruments than OTW.

There was a significant Wind X AOI interaction, F(2,18)= 5.365, p<.05 and post-hoc pairwise comparisons revealed that in high wind conditions, pilots’ visual attention OTW was higher (26.6%) than in the nominal condition (21.11%), t(9) = 4.808, p<.01. The pairwise comparisons between the nominal and high-wind condition were not significant for the PFD or Nav Display.

The Scenario X Wind X AOI interaction approached significance, F(2,18) = 3.073, p<.1[6]. Pairwise comparisons were conducted to evaluate the differences between the nominal condition and high wind condition, for each scenario context. In the VCSPA 800' condition, compared to the nominal condition, the high wind manipulation resulted primarily in increased visual attention OTW, t(9)=3.543, p<.01 (and to the PFD, though this difference was not significant (t(9)=1.574, p>.05) at the cost of attending to the Nav Display (t(9)=1.974, p<.1). This suggests that in the high workload condition, the PF may have less visual resources available to monitor the paired traffic and the wake on the Nav Display.

In the VCSPA 200' condition (right), the results of the high-wind scenario revealed an increase in visual attention (PDT) OTW (t(8)=2,6, p<.05 compared to the nominal model, however there was little cost to visual attention on the PFD (t(9)=.46, p>.05) or Nav Display (t(9)=1.187, p>.05. This suggests that, in the VCSPA 200' condition, pilots effectively redirected their visual attention away from less-important displays (such as the MCP, datalink, and FMS).

SCENARIO 2: FMA ERROR

This scenario evaluates the time required for a pilot to detect an incorrect mode on the flight mode annunciator (FMA), located along the top of the PFD. The FMA indicates which automation mode is engaged. In VCSPA, the FMA indicates whether the autopilots of the paired aircraft are 'coupled' or slaved. This coupling happens 12 nm from the runway threshold (about 3,000' to 4,000' altitude). Once the pilot engages the coupling automation, the FMA shows "C-SPD" C-LNAV" and "C-VNAV". This off-nominal scenario evaluates the modeled pilots' ability to detect a change in the FMA. Once the FMA error is detected, the pilot should break out of the VCSPA pairing by pressing the TOGA button. Detection of the FMA error is evaluated at three altitudes: 1000', 700', and 500'[7].

---

[6] At this early stage in concept design, a more liberal alpha of .1 is adopted throughout.

[7] Note that in VCSPA 800' at 500' the pilots have already decoupled so that scenario represents the pilots ability to detect any change on the FMA.

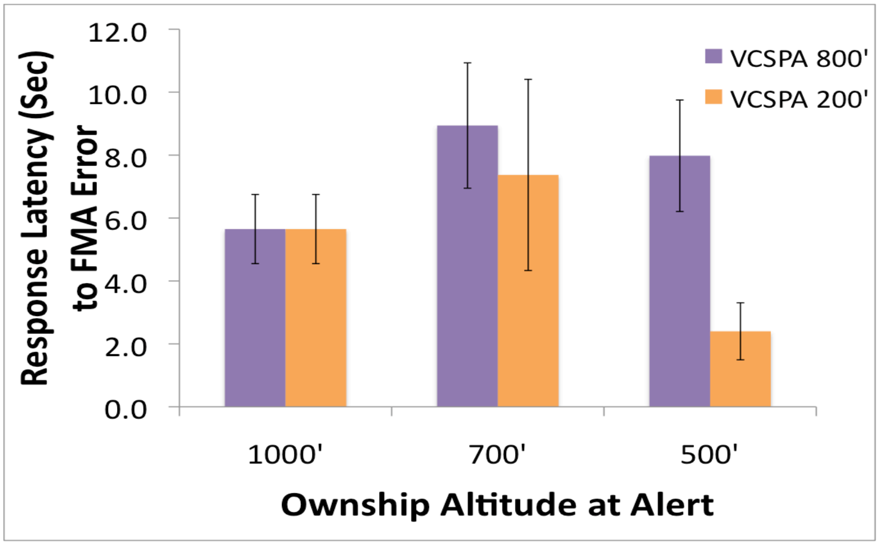


Figure 18. Pilot Flying Response Latencies (sec) to Detect Flight Mode Annunciator Errors that Occur When Ownship is at 1000', 700', and 500'. Graph presents the mean latencies of 10 monte carlo simulation runs with +/- 1 standard error.

Figure 18 presents the mean response latency for the PF to notice the FMA error and press the TOGA button to initiate the go-around procedure. A 2 (Scenario) X 3 (alert altitude) repeated measures ANOVA revealed a significant main effect for Scenario, $F(1,9) = 3.57$, $p<.1$. The response latencies were shorter in the VCSPA 200’ condition ($M=5.14$), than the VCSPA 800’ ($M=7.523$) condition.

Although the Scenario X Altitude interaction was not significant, $F(2,18) = 2.14$, $p>.05$, one can see in Figure 18 that at 10,000’, there is no difference in FMA error detection times between the VCSPA 800’ and VCSPA 200’ scenarios. This is expected, since at 10,000’, the pilots in both scenarios are still flying on full automation, and have not yet broken out of the clouds. At both 700’ and 500’ pilots are faster to detect the FMA alert in the VCSPA 200’ condition than the VCSPA 800’ condition. This occurs because, in the VCSPA 800’ condition, the pilots have started to transition their eyes OTW, as they have broken through the clouds, and are assuming manual control of the aircraft, where as the pilots in the VCSPA 200’ condition are still above the cloud ceiling and on autopilot. Their visual scan remains inside the cockpit, and thus the pilots are faster to detect the FMA error.

SCENARIO 3: RNP LOSS

This scenario evaluates the modeled-pilots’ ability to detect a text alert on the EICAS. One example of an alert is an RNP alert. The VCSPA procedure will require a minimum RNP or pilots will be expected to discontinue the VCSPA procedure. Other status messages and alerts are presented on the Upper EICAS as well. The RNAV, VCSPA800’, and VCSPA200’ models were modified to include an alert on the EICAS when the aircraft is at the following altitudes: 3,000’ and 400’.

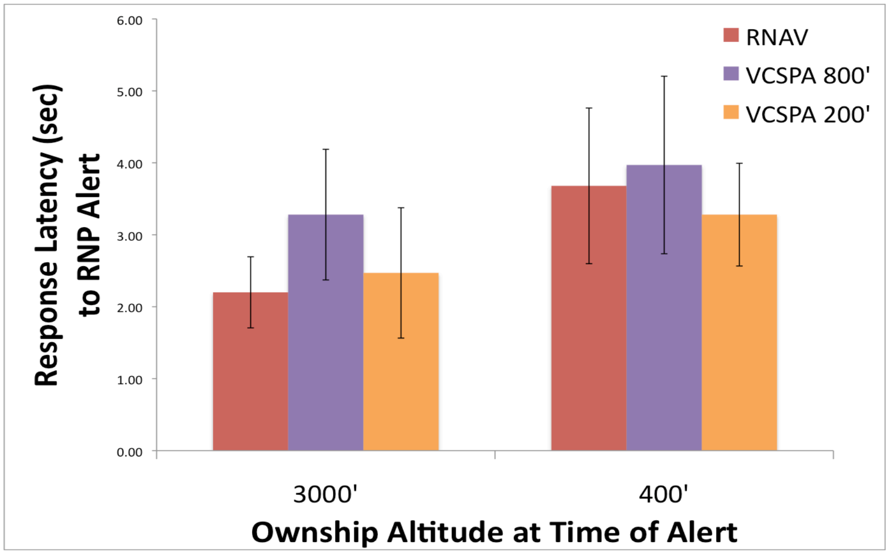


Figure 19. Pilot Not Flying Response Latencies (sec) to Detect RNP Alerts on EICAS when Ownship was at 3,000’ or 400’. Graph presents the mean latency of 10 monte carlo runs with +/- 1 standard error.

Figure 19 show the latency from onset of the RNP alert on the EICAS to first detection of the alert by the PNF. The PNF latency times are evaluated because it is assumed that the PNF may be more likely to be managing communications and system status messages than the PF. It is important to note that although differences among conditions will be discussed below, these differences failed to reach statistical significance, likely due to the high variability in latency response times exhibited in the model.

As can be seen, when the alert was presented at 3,000’, the pilot response latencies were slower in both VCSPA 800’ and VCSPA 200’ than in the RNAV condition. At this point in the scenario, pilots were initiating the pairing procedure, including receiving, reading, and accepting datalink and initiating speed-coupling automation. These tasks directed visual attention to the datalink display and the PFD, instead of system status information on the Upper EICAS, thus slowing detection of the MCP alert. This may be suggestive of a workload management problem, as identified by the focus group pilots (see Section III) and should be examined in future research.

When the alert was presented at 400’, pilots detected it slightly faster in the VCSPA 200’ condition, than in either the RNAV or VCPSA 800’ condition. In the latter two conditions, pilots’ had broken through the clouds and were manually flying the aircraft, where as in the VCSPA 200’ condition, pilots were still above the cloud ceiling and monitoring the automation, with eyes mostly inside the cockpit.

SCENARIO 4: TRAFFIC

This scenario simulates an event in which traffic taxies onto the runway while the ownship is in the land phase of flight (either when ownship is 500’ or 150’ above runway threshold) and the pilots must detect the aircraft on the runway and initiate a TOGA. The data presented in Figure 20 represent the time required for the PF to detect the aircraft on the runway and initiate the TOGA procedure. Note that the data presented are the mean of seven monte carlo trials, because in three trials the PF failed to detect the aircraft on the runway in time to initiate the TOGA procedure.

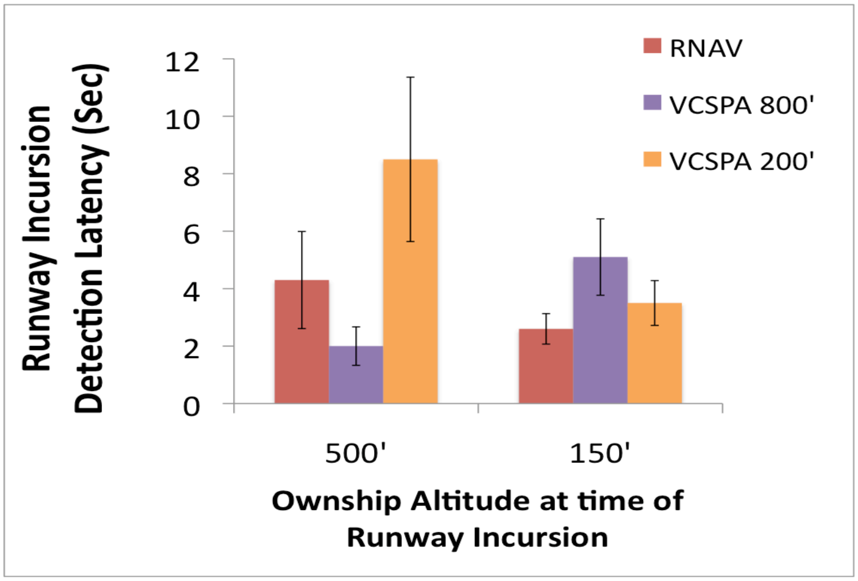


Figure 20. Pilot Flying Detection Latencies (sec) to Detect Aircraft on the Runway when Ownship is at 500' or 150'. Graph presents the mean latency of 7 monte carlo runs with +/- 1 standard error.

A Scenario (3) X Altitude (2) repeated measures ANOVA revealed a Scenario X Altitude interaction, F(2,12) = 4.33, p<.05. When the runway incursion occurred when ownship was 500' above ground (Figure 20, left side), it can be seen that, as expected, the detection time was much slower in the VCSPA 200' condition, than the VCSPA 800' condition (t(6)=2.054, p<.1. The detection time in VCSPA 200' was slower than RNAV, though the difference was not significant, t(6)=1.137, p>.05. This is expected because, at this altitude, the aircraft still has not broken through the cloud ceiling in the VCSPA 200' scenario, and the traffic on the runway is not visible to the pilot until they break through the clouds at 200' approximately 22 seconds later. Also, the 500' altitude scenario reveals that pilots were faster to detect the runway incursion in the VCSPA 800' condition than the RNAV condition, though the difference was not significant, t(6)=1.268, p> .05. This is presumably because in the VCSPA 800' condition, the pilots were monitoring traffic out the window more closely because of the closely paired traffic, then in the RNAV condition.

Implications:

1. The finding from the weather off-nominal scenario that showed that pilots attend OTW more during high-wind conditions, at the cost of attending to the NavDisplay, provides support for placing wake information on a head-up display (HUD) as well as the Nav Display. Locating wake information on the HUD, may enable the PF to maintain the priority of the primary aviation tasks

while also better managing tasks associated with separation from the lead aircraft in the CSPO environment.

2. The pattern from the weather and FMA error off-nominal scenarios suggests that since the PF will be eyes-out during CSPO, there may be a need for a different distribution of roles and responsibilities between crewmembers in the NextGen cockpit. For example, these results suggest that the PNF may need to focus more on monitoring the Wake Displays for potential transgressions of the paired traffic, freeing the PF's resources to manage the tasks associated with landing the aircraft.
3. The results of the weather, RNP alert, and FMA error off-nominal scenarios, suggested that the VCSPA 200, with autoland automation, enabled the pilots to better focus their visual attention to the cockpit displays, thus presumably affording improved awareness and safety.
4. The RNP alert scenario showed that the coupling procedure at 3,000' has the potential to distract pilots from other tasks. Future research should consider coupling implementations that minimize the load on pilots' visual attention. Caution must be taken in the design of NextGen systems to ensure the process for accessing, accepting, and loading the datalink and engaging automation are safe.

Summary

In summary, the methodical and comprehensive model validation effort illustrates a candidate process to develop, validate, and extend HPMs to predict the effect that changes to operator roles and responsibilities, candidate display technologies and automation might have on the human-system operations characteristic of in the NextGen. The credibility of the model is vastly improved when a rigorous validation effort is followed that includes formal validation of the input parameters as well as the output parameters. Such a process was completed in the phase of the research reported in the present document. Credible HPMs are in a position to be used to predict more significant operational modifications such as changes to the roles and responsibilities of the operators in an environment or changes to the candidate displays and the associated information content of the displays.

Next Steps

Future phases of the research will require model augmentations in terms of the roles and responsibilities for the cockpit crew. As documented in both Section III (Extending the Validated RNAV Model to CSPO) and Section IV (What-if Off-nominals), the PF/PNF may benefit from a change in their roles depending on the specific phase of flight. The visual fixation predictions reported in the current phase of the research could be indicative of a vulnerable portion of the approach and land operations in the NextGen. The model predicts that the operator's attention will be spent in the cockpit to a greater extent than outside of the cockpit, which in turn may result in safety implications. For instance, the more time that the pilot spends inside the cockpit, the less likely it is that they will be able to detect an unexpected event in the world – such as aircraft of the runway. This excessive load may be a candidate place for an alternate PF / PNF role –

one where the PNF take responsibility for all tasks external to the cockpit because the PF is absorbed internally. Alternatively, research may suggest that the VCSPA procedures require advanced automation to support auto-land capability (or at least lower DH) and pilot procedures should prohibit manual approaches in VCSPA conditions. However, it is important to note, that these model results do no account for known, and potentially negative, effects associated with increased automation such as inappropriate trust in automation, complacency and operator-out-of-the-loop phenomenon which may introduce other safety considerations. Further work should be conducted to define and assess alternative pilot roles and responsibilities to determine the effect on pilot workload, situation awareness, and visual attention. In addition, further work should be conducted to define pilot interface options for the presentation of separation and wake information and evaluate the effect of various information requirements on pilot workload, situation awareness, and visual attention.

V. References

Alexander, A.L., & Wickens, C.D. (2005, Retrieved March 7, 2008). *Synthetic vision systems: Flightpath tracking, situation awareness and visual scanning in an integrated hazard display*. Paper presented at the Proceedings of the 13th International Symposium on Aviation Psychology (CD-ROM), Oklahoma City, OK.

Anders, G. (2001). *Pilot's Attention Allocation during Approach and Landing: Eye- and Head-Tracking Research in an A330 Full Flight Simulator*. Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH.

Arditi, A. , & Azueta, S. (1992). *Visualization of 2-D and 3-D aspects of human binocular vision*. Paper presented at the Society for Information Display International Symposium.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation*. London: Academic Press.

Badler, N.I., Phillips, C.B., & Webber, B.L. (1993). *Simulating humans: Computer graphics, animation, and control*. Oxford: Oxford University Press.

Campbell, G.E., & Bolton, A.E. (2005). HBR validation: interpreting lessons learned from multiple academic disciplines, applied communities, and the AMBR project. In K.A. Gluck & R.W. Pew (Eds.), *Modeling human behavior with integrated cognitive architectures: Comparison, evaluation and validation* (pp. 365-395). New Jersey: Lawrence Erlbaum & Associates.

Card, S.K., Moran, T.P., & Newell, A. (1983). *The psychology of human computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.

Dixon, S.R., & Wickens, C.D. (2003). *Imperfect Automation in Unmanned Aerial Vehicle Flight Control*. Boulder, CO: Micro Analysis and Design Operations.

Elkind, J., Card, S.K., Hochberg, J., & Huey, B.M. (1989). *Human performance models for computer-aided engineering*. Washington, D.C.: National Academy Press.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102(2)*, 211-245.

Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381-391.

Gluck , K.A., & Pew, R.W. (2005). Modeling human behavior with integrated cognitive architectures: comparison, evaluation and validation. In K.A. Gluck & R.W. Pew (Eds.), *Modeling human behavior with integrated cognitive architectures: comparison, evaluation and validation*. New Jersey: Lawrence Erlbaum & Associates.

Gore, B.F. (2002a). An emergent behavior model of complex human-system performance: An aviation surface related application. *VDI Bericht 1675*, 313-328.

Gore, B.F. (2002b). *An emergent behavior model of complex human–system performance: An aviation surface-related application*. Paper presented at the Fifth Annual SAE International Conference and Exposition - Digital Human Modeling for Design and Engineering.

Gore, B.F. (2008). Chapter 32: Human performance: Evaluating the cognitive aspects. In V. Duffy (Ed.), *Handbook of digital human modeling* (pp. 32:31-32:18). Boca Raton, Fla: CRC Press Inc.

Gore, B.F. (2010). Man-machine integration design and analysis system (MIDAS) v5: Augmentations, motivations, and directions for aeronautics applications. In P.C. Cacciabu, M. Hjalmdahl, A. Luedtke & C. Riccioli (Eds.), *Human Modelling in Assisted Transportation*. Heidelberg: Spinger.

Gore, B.F., & Corker, K. (1999). *System interaction in free flight: A modeling tool cross comparison*. Paper presented at the International Conference on Digital Human Modeling for Engineering and Design, The Hague, Netherlands.

Gore, B.F., Hooey, B.L., Wickens, C.D., & Scott-Nash, S. (2009, July19-24). *A computational implementation of a human attention guiding mechanism in MIDAS v5*. Paper presented at the 12th International Conference, HCI International 2009, San Diego, CA.

Gore, B.F., Hooey, B.L., Wickens, C.D., Sebok, A., Hutchins, S, Salud, E., et al. (2009). *Identification of NextGen Air Traffic Control and Pilot Performance Parameters for Human Performance Model Development in the Transitional Airspace* (No. NASA Final Report: NRA #NNX08AE87A). San Jose, CA: San Jose State University.

Gugerty, L. (1998). Evidence from a partial report task for forgetting in dynamic spatial memory. *Human Factors, 40*(3), 498-508.

Hamilton, D.B., Bierbaum, C.R. , & Fullford, L.A. (1990). *Task analysis/workload (TAWL) user's guide, version 4 (Research Project 91-11)*. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences (AD A241 861).

Harber, R.N., & Hershenson, M. (1980). *The Psychology of Visual Perception* (Vol. 360). London: Holt, Reinhart, & Winston.

Hooey, B.L., & Foyle, D.C. (2003). *A Post-Hoc Analysis of Navigation Errors During Surface Operations: Identification of Contributing Factors and Mitigating Solutions*. Paper presented at the AIAA's 3rd Aviation Technology, Integration, and Operations (ATIO) Technical Forum.

Hooey, B.L., & Foyle, D.C. (2008). Advancing the state of the art of human performance models to improve aviation safety. In B.L. Hooey & D.C. Foyle (Eds.), *Human performance modeling in aviation* (pp. 321-349). Boca Raton: CRC Press/Taylor & Francis.

Hüttig, G., Anders, G., & Tautz, A. (1999). *Mode Awareness in a modern Glass Cockpit – Attention Allocation to Mode Information*. Paper presented at the 10th International Symposium on Aviation Psychology, Columbus, OH.

JPDO, Joint Planning and Development Office (2009). Concept of operations for the next generation air transportation system. In Joint Planning and Development Office (Ed.) (Vol. 3). Washington, DC: JPDO.

Keller, J.W., Leiden, K., & Small, R. (2003). Cognitive task abnalysis of commercial jet aircraft pilots during instrument approaches for baseline and synthetic visision displays. In D.C Foyle, A. Goodman & B.L. Hooey (Eds.), *Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays (NASA/CP-2003-212267)*. Moffett Field, CA: NASA Ames Research Center.

Lubin, J. , & Bergen, J. (1992). NASA cockpit display visibility modeling project. Moffett Field, CA: SRI/David Sarnoff Research Center.

McCracken, J.H., & Aldrich, T.B. (1984). *Analysis of selected LHX mission functions: Implications for operator workload and system automation goals*. Fort Rucker, AL: Anacapa Sciences, Inc.

Mitchell, D. K. (2000). *Mental workload and ARL workload modeling tools*. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Mumaw, R.J., Sarter, N., & Wickens, C.D. (2001). *Analysis of pilots' monitoring and performance on an automated flight deck*. Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39*(2), 230-253.

Robinson, G.H. (1979). Dynamics of the Eye and Head During Movement between Displays: A Qualitative and Quantitative Guide for Designers. *Human Factors*(21), 343-352.

Robinson, G.H., Koth, B.W., & Ringenbach, J.P. (1976). Dynamics of the Eye and Head During an Element of the Visual Search. *Ergonomics, 19*, 691-709.

Shannon, C.E., & Weaver, W. (1949). *Mathematical theory of cognition*. Urbana and Chicago: University of Illinois.

Verma, S., Lozito, S., & Trott, G. (2008). *Preliminary Guidelines on Flight Deck Procedures for Very Closely Spaced Parallel Approaches*. Paper presented at the International Council for the Aeronautical Sciences (ICAS) 2008 Congress.

Welford, A.T. (1976). *Skilled performance*. Glenview, Il: Scott Foresman.

Wickens, C.D., Goh, J., Helleberg, J., Horrey, W., & Talleur, D.A. (2003). Attentional Models of Multi-Task Pilot Performance using Advanced Display Technology. *Human Factors, 45*(3), 360-380.

Wickens, C.D., & McCarley, J.S. (2008). *Applied Attention Theory*. Boca Raton, Fla: CRC Press, Taylor and Francis Group.