NASA's Use of Human Performance Models for NextGen Concept Development and Evaluations

Brian F. Gore[1], Becky L. Hooey[1], and David C. Foyle[2]
[1]San Jose State University/NASA Ames Research Center
[2] NASA Ames Research Center
MS 262-4, PO Box 1
Moffett Field, CA 94035-0001
650-604-2542, Brian.F.Gore@nasa.gov

Keywords:
NASA, Human Performance Model, NextGen Models, MIDAS v5 Validation

ABSTRACT: Integrated Human Performance Model (HPM) validity is a paramount concern when HPM predictions are used for next-generation aviation system development. HPM validity is a challenge because of the integrated nature of the HPM and because many of the embedded behaviors may not be readily observed. A rigorous validation process is required to arrive at valid integrated HPMs and improve the credibility of the models being developed. This credibility impacts the subsequent use of the model to explore concepts being proposed for future systems. The current paper will highlight a recent methodical validation approach that was developed and applied to a Federal Aviation Authority (FAA)-National Aeronautics and Space Administration (NASA) HPM of a candidate NextGen concept of operations using the Man-machine Integration Design and Analysis System (MIDAS v5). The HPM that was developed was deemed valid from multiple levels using multiple input and output parameters.

1. Introduction

Modeling and simulation (M&S) techniques, particularly human behavior models, play an important role when complex human-system concepts are being proposed, developed, and tested across many of the ten NASA centers. For instance, NASA Johnson Space Center (JSC) utilizes M&S to represent environments, physical structures and equipment components, crew stations, planets and planetary motions, gravitational effects, illumination, human anthropometric and biomechanics, among a host of other domains. NASA Ames Research Center (ARC) also possesses a number of M&S capabilities ranging from airflow, flight path models (e.g., Airspace Concept Evaluation System, -ACES), aircraft models, scheduling models (e.g., Core-XPRT, Science Planning InterFace to engineering - SPIFe), human performance models (HPMs), and bioinformatics models, among many other kinds of M&S capabilities. A comprehensive validation process of one of the many NASA M&S capabilities, an ARC-related HPM capability termed the Man-Machine Integration Design and Analysis System (MIDAS) is highlighted in the current paper because of its relevance to the field of integrated human behavior representation validation.

1.1 Complex Systems, Concept Development and Testing, Human Performance Models (HPMs)

Complex systems are those that include human operators interacting with technology and automation, to carry out multiple interacting, and often conflicting, tasks. In such complex systems, one of the advantages of a HPM is the ability to model critical events that cannot be studied fully with HITL subjects, due to safety concerns, cost considerations, or practical difficulties associated with the simulation of very rare events. Another advantage of the HPM methodology is that estimates of human-system interaction can be made on systems for which concepts, technologies, and automation cannot be tested in empirical simulations because the system concepts are too new, too difficult, or too dangerous for the human operator (e.g. space operations, technologies for advanced aircraft system concepts, advanced medical technologies, and advanced surface transportation technologies).

1.2 Objectives

Recent research at NASA Ames Research Center aimed to: Develop valid HPMs of aircraft approach and land operations; Use these models to evaluate the impact of proposed NextGen concepts for Closely Spaced Parallel Operations (CSPO) on pilot performance; and, Draw conclusions regarding flight deck displays and pilot roles and responsibilities for NextGen CSPO concepts. The approach was to first develop and validate a model of current-day area navigation (RNAV) operations, extend this validated model to NextGen CSPO concepts (Verma, et al., 2008), and then conduct "what-if" simulations to evaluate the predicted effect of concept designs on the human operator's performance.

1.3 The Man-machine Integration Design and Analysis System Version 5 (MIDAS v5)

MIDAS v5 is a dynamic, integrated human performance modeling and simulation environment that facilitates the design, visualization, and computational evaluation of complex man-machine system concepts in simulated operational environments (Gore, 2008). 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. HPMs like MIDAS provide a flexible and economical way to manipulate aspects of the operator, automation, and task environment for simulation analyses (Gore, 2008; Gore, Hooey, Foyle, & Scott-Nash, 2008; Hooey & Foyle, 2008a).

1.4 MIDAS v5 Model of Current-day Approach and Land Operations

Using NASA's MIDAS v5, a high-fidelity model of a two-pilot commercial transport crew flying current-day RNAV approach and land operations was developed. The model, containing over 970 tasks, was based on cognitive task analyses and cognitive walkthroughs conducted with commercial pilots and air traffic controllers.

1.5 MIDAS Approach and Land Validation Effort

The current-day RNAV model was validated using a methodical, multi-dimensional approach (as illustrated in Figure 1) that validate the model's inputs, process models, and outputs.

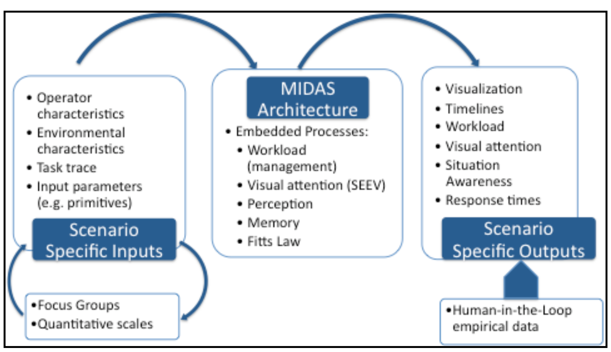

Figure 1: Methodical Validation Approach Used for the MIDAS v5 Approach and Land Model.

1.5.1 Input Validation

The model inputs were validated using focus group sessions comprised of 8 commercial transport 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. Validating the model inputs included two aspects. First, a formal analysis of the task trace was conducted to determine the extent to which the modeled tasks represent the pilots' actual tasks (see Gore, et al., 2011). 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. The pilots 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.

1.5.2 Process Model Validation

The MIDAS process models have also been validated. The MIDAS processes 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, workload, and situation awareness (SA). The process models contained within MIDAS have all previously been validated in different application domains. The operator attention model is based on the salience, expectancy, effort and value (SEEV) model of information retrieval (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). The perception model was validated with empirical data on visual detection (Arditi & Azueta, 1992; Harber & Hershenson, 1980; Lubin & Bergen, 1992). The three-stage memory model was validated by Ericsson and Kintsch, (1995), the situation awareness model by Hooey, et al., (2010), and the workload process model by Gore, et al., (in process).

1.5.2 Output Validation

After the model inputs and process models were deemed to accurately represent human operator capacities, the model outputs, workload and visual attention, of the refined model were statistically compared to existing human-in-the-loop (HITL) simulation data. 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. The workload model output according to three phases of flight correlated with a comparable HITL study (Hooey & Foyle, 2008b) with r^2 of .54 for overall workload. The individual workload dimensions of three phases of flight also correlated positively with the HITL study with r^2 ranging from .55 to .94. Visual attention as measured by percent dwell time on the primary flight display (PFD), the Navigation Display (ND) and out the window (OTW) areas of interest correlated with three HITL studies (r^2 = .99). These validation results provide confidence that the model validly represents pilot performance.

1.6 Extending to NextGen Concepts

The validated baseline RNAV model was then extended to represent a NextGen CSPO concept termed Very Closely Spaced Parallel Operations (VCSPA; Verma, et al, 2008) developed and evaluated at NASA Ames Research Center. Two operational implementations of the VCSPA concept were modeled: 1) VCSPA 800', with a 800' cloud ceiling and manual flight after a decision height (DH) of 650'; and, 2) VCSPA 200' with a 200' cloud ceiling and autoland capability. The model inputs (task trace and input parameters) were validated using the same scenario-based focus groups as above. 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 ND 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. In both the VCSPA 800' and 200' conditions, the MIDAS input model was modified to reflect the focus group recommended procedural changes. Thus, the refined model more accurately represented the information being communicated as well as the task sequence for both the NextGen CSPO manipulations.

After the inputs had been rigorously validated through the focus group session, 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). 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, 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 the assumption of autoland automation. The MIDAS model predicts that both VCSPA scenarios will draw the pilots' attention to the traffic and wake information on the ND at the expense of attending to the PFD and OTW. This model implementation illustrates a potential system vulnerability if the pilot needed to respond to time critical information presented on the PFD or OTW. These were examined in the "What-if" evaluation phase discussed next.

1.7 Conducting "What-if" Manipulations

Next, with validated models of RNAV and VCSPA operations, "what-if" scenarios were conducted to explore the impact of the NextGen CSPO concept on pilot 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) were asked to identify potential off-nominal events after a scenario-based walk-through of RNAV and CSPO approaches. Second, a search of the Aviation Safety Reporting System (ASRS) database yielded 199 incident reports, which was narrowed down to 13 potential off-nominal events. Third, findings from previous research (Gore, et al., 2009) were adopted that included a systematic approach to identifying off-nominal events and their contributing factors, and characterizing off-nominal causal factors using a modified taxonomy: Environment, System Management, Human and Machine (Foyle & Hooey, 2003). 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 (percent dwell time, PDT), and time to detect or respond to the event were recorded across ten runs of each off-nominal model. These output were then compared to the "nominal" performance data and implications for CSPO operations including pilot roles and responsibilities and flight deck displays were generated (Gore, et al., 2011). The results from the various off nominal events provide support for including such variables as candidates to include in further HITL simulations. For instance, the high-wind off-nominal scenario showed that pilots attend OTW more during high-wind conditions, at the cost of attending to the ND. This result provides support for placing wake information on a head-up display (HUD) as well as the ND. 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. Additional results related to the other off-nominal events and their implications can be found in Gore, et al. (2011).

2. Discussion and Conclusion

This paper highlights one approach to validate a complex HPM, and the manner that the validated model can be extended to predict likely human-system issues in a future system. When new technologies and procedures are introduced into flight deck operations, as are expected to occur with the NextGen CSPO concepts (JPDO, 2009), new human-system vulnerabilities may surface. HPMs may be used to predict the points at which the human-system vulnerabilities are most likely, and identifying these vulnerabilities is useful only if the output is valid. 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 validation results of workload and visual attention provide confidence that the model validly represents pilot performance. This model was then extended to: 1) evaluate current-day and candidate NextGen operations; and, 2) to evaluate candidate off-nominal operations in a "what-if" format. Even though there are no HITL data available for the NextGen environment, it was assumed that the NextGen output were valid because the same rigorous process was utilized to create the NextGen scenario and validate the model inputs. Any differences were assumed to be due to procedural differences in the NextGen. Similarly, any differences between the off-nominal scenario and the baseline model were assumed to be the result of differences in the off-nominal procedures. The findings yield implications for candidate NextGen roles and responsibilities and flight deck displays and automation.

The methodical and comprehensive model validation effort illustrates a candidate process to develop, validate, and extend HPMs to predict the effect that NextGen changes to operator roles and responsibilities, candidate display technologies and automation might have on human-system performance. 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.

3. References

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.

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

Foyle, D.C., & Hooey, B.L. (2003). *Improving evaluation and system design through the use of off-nominal testing: A methodology for scenario development.* Paper presented at the Twelfth International Symposium on Aviation Psychology, Dayton, Ohio: Wright State University.

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., Hooey, B.L., Foyle, D.C., & Scott-Nash, S. (2008, July 14-17). *Meeting the challenge of cognitive human performance model interpretability though transparency: MIDAS v5.x.* Paper presented at the Applied Human Factors and Ergonomics International Conference, Las Vegas, Nevada.

Gore, B.F., Hooey, B.L., Socash, C., Haan, N., Mahlsted, E., Bakowski, D.L., Gacy, A.M, Wickens, C.D., Gosakan, M, & Foyle, D.C. (2011). Evaluating NextGen Closely Spaced Parallel Operations Concepts with Human Performance Models. *HCSL Technical Report (HCSL-11-01)*. Moffett Field, CA: NASA Ames Research Center.

Gore, B.F., Hooey, B.L., Wickens, C.D., Sebok, A., Hutchins, S, Salud, E., Small, R., Koenecke, C, & Bzosteck, J. (2009). *Identification of NextGen Air Traffic Control and Pilot Performance Parameters for Human Performance Model Development in the Transitional Airspace* (NASA Final Report: NRA #NNX08AE87A). San Jose, CA: San Jose State University.

Gore, B.F., Hooey, B.L., Wickens, C.D., Socash, C., Gacey, A.M., Brehon, M., Gosakan, M, & Foyle, D.C (in process). *The MIDAS workload model (HCSL-11-02)*. Moffett Field, CA: NASA Ames Research Center.

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

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

Hooey, B.L., & Foyle, D.C. (2008b). Aviation Safety Studies: Taxi Navigation Errors and Synthetic Vision System Operations. In D.C. Foyle & B.L. Hooey (Eds.), *Human performance modeling in aviation* (pp. 29-64). Boca Raton: CRC Press/Taylor & Francis.

Hooey, B.L., Gore, B.F., Wickens, C.D., Salud, E., Scott-Nash, S., Socash, C., & Foyle, D.C. (2010). *Modeling Pilot Situation Awareness*. In P.C. Cacciabu, M. Hjalmdahl, A. Luedtke & C. Riccioli (Eds.), *Human Modelling in Assisted Transportation*. Heidelberg: Spinger.

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.

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 (Technical note ASI 479-024-84[b])*. Fort Rucker, AL: Anacapa Sciences, Inc.

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.

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.

Author Biography

BRIAN GORE is a Principal Investigator for San Jose State University, and NASA's Technical Lead for the Man-machine Integration Design and Analysis System (MIDAS) at NASA Ames Research Center.

BECKY HOOEY is a Principal Investigator for San Jose State University at NASA Ames Research Center.

DAVID FOYLE is the NASA Civil Servant POC for the machine Integration Design and Analysis System (MIDAS) at NASA Ames Research Center.

Acknowledgement

The composition of this work was supported by the Federal Aviation Administration (FAA)/NASA Inter Agency Agreement DTFAWA-10-X-80005 Annex 5 (FAA POCs Tom McCloy, Dan Herschler; NASA POC David C. Foyle). Portions of this work were published as a final deliverable from the HCSL to the FAA and to the HCII 2011 conference. The authors would like to thank Connie Socash, Christopher Wickens, Marc Gacy, and Mala Gosakan from Alion Science and Technology (MAAD Operations) for their invaluable model development support, Nancy Haan (Dell/Perot Systems), Eric Mahlstedt and Deborah Bakowski (San Jose State University) for their support on the project and all reviewers for their insightful comments.