

# Development, Testing, and Validation of a Model-Based Tool to Predict Operator Responses in Unexpected Workload Transitions

Angelia Sebok, Christopher Wickens and Robert Sargent  
Alion Science and Technology, 4949 Pearl East Circle, Suite 100, Boulder, CO 80301

One human factors challenge is predicting operator performance in novel situations. Approaches such as drawing on relevant previous experience, and developing computational models to predict operator performance in complex situations, offer potential methods to address this challenge. A few concerns with modeling operator performance are that models need to be realistic, and they need to be tested empirically and validated. In addition, many existing human performance modeling tools are complex and require that an analyst gain significant experience to be able to develop models for meaningful data collection. This paper describes an effort to address these challenges by developing an easy to use model-based tool, using models that were developed from a review of existing human performance literature and targeted experimental studies, and performing an empirical validation of key model predictions.

## INTRODUCTION

This paper describes the development, testing, and empirical validation of a research and model-based tool to predict operator performance in unexpected workload transitions. The tool, called S-PRINT (Space Performance Research Integration Tool) was developed for NASA to predict and examine astronaut performance on long-duration space missions. As described previously (Sebok, Wickens, Clegg & Sargent, 2014) this tool was developed as a plug-in to IMPRINT (the Improved Performance Research Integration Tool; Allender, 2000), and provides a number of enhancements and new capabilities within IMPRINT.

Computational modeling is a technique to predict human performance in novel, not-yet-tested and not-easily-tested situations, such as long-duration space exploration missions. One of the primary advantages of modeling is that it allows analysts to evaluate different situations. Specifically, users can tweak different parameters – e.g., of the situation in which the personnel will be working, the design of the equipment they will be using, the number of team members, or the fatigue the operators could experience – to identify the effects of these factors on predicted performance. Such analyses can be used to identify the need for different equipment designs, staffing configurations, training, task design, or scheduling.

There are challenges associated with modeling. First, there needs to be a solid, defensible body of work supporting the models developed. Another challenge is that models need to be validated. They need to be tested against actual operator performance in a real or simulated situation that matches the model (Wickens & Sebok, 2014). Finally, tool usability is a concern. Many different personnel could benefit from being able to use models to predict operator performance in novel situations. But modeling tools (because of the many capabilities they offer and the complexity of what users are attempting to do) typically require that the analyst have training and significant experience before being able to develop intricate models of human performance.

In this project, we developed a model-based tool for NASA personnel involved in planning and monitoring long-duration space missions. S-PRINT is intended to be used by human performance specialists, mission planners, automation designers, and possibly even astronauts. All of these personnel

have an interest in astronaut performance in space missions, from different perspectives. Given the variety of users and the different tasks they will likely perform, we emphasized the need for an easy to use tool.

S-PRINT is based on a particular worst-case scenario. It investigates operator performance in a workload transition due to an automation failure. The tool includes a set of component models, described below, that were developed to predict the effects of relevant factors on human performance. Each of these component models were developed based on empirical, human-in-the-loop (HITL) data.

## APPROACH

Our approach to this work was first to identify key factors likely to affect human performance in long-duration space missions: fatigue effects on performance, human automation interaction, and task overload performance. We developed algorithms (or component models) to simulate these factors. We also developed task network models of relevant scenarios that astronauts might encounter on a space mission.

As decades of experience with human spaceflight indicate, astronauts experience poor sleep in space compared with their sleep on earth (Whitmire et al., 2009). Further, on long-duration missions, there will probably be many routine tasks performed day in and day out. Monotony is a possible (but debated) concern. The conditions of poor-sleep-induced fatigue and routine tasks can lead to operator complacency in monitoring and interacting with automation.

A second concerning situation is a potential automation failure, described in more detail below. In our research, the automation failure causes a transition to a much higher workload, the third factor. The situation we evaluate, because of its “worst case” nature, is one in which a fatigued astronaut is unexpectedly confronted with an automation failure that requires the operator to select certain tasks and shed others.

## Component Models

*Fatigue Models.* An integrated fatigue model (Wickens, Laux, Hutchins, & Sebok, 2014) was developed based in part on an extensive meta-analysis of research (Wickens, Hutchins, Laux & Sebok, 2015) that examined the effects of four fatigue

conditions: sleep deprivation, sleep restriction, circadian cycle and sleep inertia effects on operator performance on *complex tasks*. Operator performance is impacted in the task network model through performance shaping factors (PSFs) that predict longer task completion times or diminished task accuracy in fatigued conditions. PSF effects are readily modeled and implemented in IMPRINT.

Investigating fatigue effects on complex tasks is novel, as much of the existing fatigue literature and models are based on simple task performance, such as the psychomotor vigilance task (PVT). Interestingly, our findings revealed that complex task performance is generally less impacted by fatigue than is simple task performance (Wickens, Hutchins et al., 2015), although the impacts do remain substantial. The multitasking, challenging nature of complex tasks provides stimulation that keep the operator more aroused and engaged in the work.

Our research also included development of a sleep inertia model. This algorithm was also based on data gathered in multiple experimental studies. We found that sleep inertia, or the grogginess upon first awakening, results in a significant drop (a 35 percent decrement) in performance compared with baseline, rested performance. This decrement is worse if the operator has been sleep deprived or sleep restricted, is awakened during circadian night, or is awakened from deep sleep (2-6 hours in duration). A recovery period occurs after awakening. From the empirical literature, we approximate the half-life (where performance is recovered half way back to baseline) to be 15 minutes, and full recovery to occur at 30 minutes after awakening.

In addition to our custom algorithms developed for S-PRINT, IMPRINT also includes fatigue models (e.g., the Sleep, Activity, Fatigue and Task Effectiveness [SAFTE] plug in; Hursch, 2003) that address sleep-related fatigue effects on *simple* task performance. Because S-PRINT was built within IMPRINT, we use these existing SAFTE algorithms to assess performance time for simple tasks.

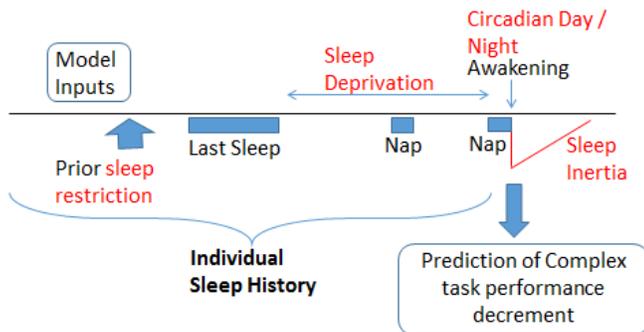


Figure 1: The Analytic Sleep Model

Figure 1 depicts how the sleep history of an individual, prior to the time at which performance is to be predicted, can be distilled into the four different components of our fatigue model: sleep restriction, sleep deprivation, circadian cycle effects, and sleep inertia. Different PSFs are applied to individual task performance depending on the astronaut’s sleep history and whether the task is simple or complex.

*Human-Automation Interaction Models.* The goal of the model of human-automation interaction (HAI) was to develop a robust, user-friendly model that could be applied across many different types of automation designs.

In developing the HAI model we considered three generic properties of an automation system (listed below) that satisfied the following criteria: *Sufficient data existed* on each property to make reasonable predictions about effects on human performance. *Aspects of the property could be quantified* in order to make quantitative predictions on human-system performance. The property was *sufficiently understandable* that users of S-PRINT could assign values to the property in the tool without the need for excessive expertise in HAI. In brief, the three properties are:

1) Degree of Automation. A prior meta-analysis (Onnasch et al., 2013) evaluated the effects of automation implementation on human and system performance. The research considered stages and levels (referred to collectively as “degree”) of automation, where stages progress from information filtering or alerting, to information integration and diagnosis, to decision making, to action or control. The levels within each state range from low to high automation authority (Parasuraman, Sheridan & Wickens, 2000). Table 1 represents these two dimensions.

Table 1: Stages (columns) and Levels (rows) of Automation

	Alerting	Diagnosis	Decision Making	Control
High	Indicates status	Interprets the condition	Executes the action unless the operator vetoes it	All actions accomplished automatically
Low		Provides integrated data	Recommends an action the operator can select	Some actions accomplished manually
Manual	Operator seeks data	Operator interprets the data	Operator decides on a course of action	Operator implements the action

The meta-analysis found that system performance improves as degree of automation increases, but on the infrequent occasions when automation fails, higher degrees of automation produce more serious consequences. Our model quantifies the **degree of automation**, by combining the impact of higher levels and later stages.

2. Automation Reliability. Increasing reliability of **imperfect** automation has analogous effects to increasing degree of automation. Higher reliability produces better performance when automation works as intended, but more problematic outcomes occur on the (increasingly rare) occasions when it fails. Another meta-analysis of the literature (Wickens, Hooey, et al., 2009) documents the much greater performance cost of the failure of highly reliable automation.

3. Alert Absence Penalty. In cases where automation fails, we apply an additional penalty when there is no alerting of the failure. Such a penalty would be applied if a display continues to portray data, even if those data are bad because of a failed sensor. On the other hand, if the display goes blank when the sensor fails, or if the screen contains a large X over it, the penalty would not be applied, as such indications are salient. This penalty is partly based on research (Wickens, Clegg, Vieane & Sebok, 2015) that investigated automation bias and complacency in unexpected failures of an automated guidance

system. This research found that operators showed a significant bias to follow incorrect yet plausible guidance.

Together all three of these properties (automation degree, reliability, and alert absence) are assigned numerical values depending on their potential penalty when automation fails. These values are summed in an equally weighted fashion, to provide an overall cost or performance decrement (a PSF that increases task times) when an automation failure occurs.

**Implementation within IMPRINT and S-PRINT.** The HAI model was implemented in two primary ways in S-PRINT, the first of which was described above, using a PSF derived from characteristics of the automated system. The second way the HAI model was implemented in S-PRINT requires coding on the part of the IMPRINT modeler. The model developer identifies the specific types of automation to be modeled, and characterizes operator tasks for each of the different modes of operation. These automation modes typically have different implications for operator interaction (e.g., task allocation, control actions, information presentation, feedback), so the specific tasks associated with different modes must be modeled explicitly. The model developer assigns the appropriate automation stages and levels (Table 1) to the different modes of automation. Further, the modeler specifies the tasks that are associated with different automation failure types. S-PRINT, as provided to NASA, includes two task network models (or scenarios) that have already been developed using the HAI coding techniques.

The S-PRINT (novice) user loads a model, selects the automation mode to investigate, and specifies the reliability of that mode for different possible failure types. The S-PRINT user decides whether or not to run the model with a failure, and – if a failure is to occur – identifies the specific failure and the salience of the indication. The S-PRINT HAI PSF model is then invoked to calculate and apply appropriate penalties to operator performance based on the failure of highly-automated, highly reliable systems, particularly when those failures are of low or no salience.

**Strategic Task Overload Management Model.** The third component model addresses operator task selection behavior when the operator is overloaded. The Strategic Task Overload Model (STOM; Gutzwiller, Wickens, Santamaria, 2015; Wickens, Vieane, Clegg, Sebok & Janes, 2015) predicts operator task selection based on four task factors (Gutzwiller et al, 2014). The *interest or engagement* the operator has in the task, the *priority* of the task, the *difficulty* of the task, and the *salience* of the task all affect the likelihood that an operator will choose to stay with an on-going task or switch to a potential alternate task. If the decision is to switch, STOM predicts which alternate task is the most likely to be selected.

### **Task Network Models / Scenarios**

In addition to the component models, S-PRINT includes two IMPRINT task network model scenarios that simulate an astronaut performing complex tasks with different automated systems. The tool provides output files that give distribution data on predicted task performance. The primary task network model for this project addresses a possible scenario in a long-duration space exploration mission. An astronaut is using a

robotic arm to transport another astronaut, while also monitoring an environmental process control system that maintains the atmospheric conditions in the spacecraft. While current day space operations include considerable support from ground personnel (e.g., monitoring and controlling the atmospheric process control system on the International Space Station), lengthy communications delays expected in long duration missions will require that control functions currently managed by ground personnel will need to be taken over by astronauts or by automation.

Both automated systems in the primary IMPRINT model (robotic arm and process control) can be used in either manual or automated mode. The manual mode of the robotic system requires the operator to use joysticks to manipulate the arm through a trajectory. In the automated mode, the robotic system executes the trajectory, and the operator monitors progress and intervenes if needed. The process control system indicates when a fault or deviation occurs. The manual mode requires that the operator diagnose and manage the problem. In the automated mode, the process control system provides specific diagnosis and repair guidance.

Even without any automation failures, the baseline scenario model introduces a process control situation requiring astronaut intervention. Different automation failures can occur. The robotic arm can require the operator to bypass an unexpected obstruction. The automated mode of the robotic arm control can fail to work properly, necessitating operator intervention and a return to manual mode. The decision aid for the process control system can fail to give guidance, or it can give an incorrect diagnosis and incorrect guidance.

In addition to the robotic arm - process control scenario, S-PRINT includes a model to evaluate and compare operator performance using one of three fire detection and suppression systems. The least automated system is a simple smoke detector that sounds when it senses smoke. An intermediate system provides an annunciation, but it also includes a digital map that indicates the location of the triggered detector. The most highly automated system includes the detector and locator map, and it includes a sprinkler that will put out the fire. Automation failures can occur, including, e.g., failure to detect the fire, indication of an incorrect location, or failure to suppress the fire.

## **TOOL DEVELOPMENT**

Two main goals with S-PRINT were 1) to provide a simple (novice mode) interface to allow a more diverse group of users (not only human performance modeling experts) to use the tool to obtain meaningful results, and 2) to include the capabilities to allow expert users (modeling experts) to develop a variety of relevant scenarios. The intent is that NASA (or any other future user) will be able to have internal experts or consultants develop new scenarios as needed, while many different users could use the tool to investigate particular situations in which they have interest.

S-PRINT includes a graphical user interface, shown in Figure 2, that provides users with access to the options they can investigate. They select the scenario of interest (e.g., the robotic arm and process control scenario or a fire detection

and suppression system, described above), and then specify characteristics of the situation they want to investigate. S-PRINT runs the underlying IMPRINT task network model to generate data on predicted operator performance. Table 2 identifies the different types of analyses that S-PRINT users can perform. The first column identifies the topic of interest, the second column lists the inputs that the user provides to the interface, and the third column indicates the relevant result files for answering the questions of interest.

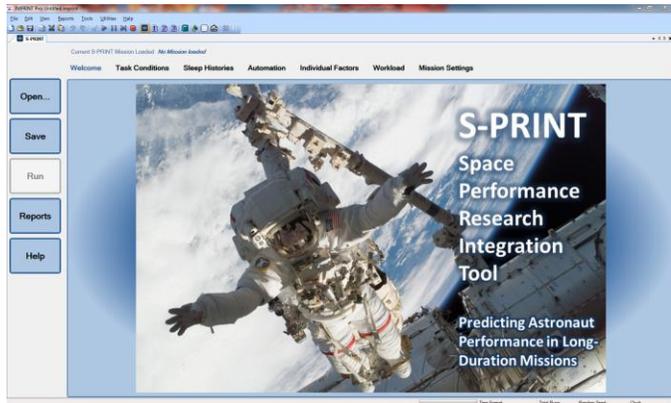


Figure 2: S-PRINT Interface

Table 2: S-PRINT Interface – Topics, Inputs, and Outputs

What S-PRINT Evaluates	What the User Enters	Relevant Reports
Operator and system performance with automated systems	Select a library model	Mission completion time and success Operator workload Task timeline
Fatigue	Astronaut sleep histories Awake or asleep at model start	Fatigue impacts on task time and accuracy Mission completion time and success
Automation	Mode Reliability Failure type Salience of failure indication	Mission completion time and success Tasks that were performed / deferred Operator workload
Task Overload Management	Task factor ratings for each task Priority Difficulty Interest Salience	Tasks that were performed / deferred Mission completion time and success Task timeline
Operator Factors	Use of protective clothing Individual fatigue resistance Workload threshold Task management strategies	Mission completion time and success Tasks that were performed / deferred Operator workload Task timeline
Workload	Adjust the components of workload for individual tasks	Operator workload Mission completion time and success

The component models for fatigue, automation, and task overload management were described previously. In addition, S-PRINT allows users to evaluate the effects of Operator

Factors and individual task Workload on predicted performance. These two capabilities are standard IMPRINT features that were leveraged in this project. The effects of protective clothing (such as the use of fully-enclosed, helmet, gloves, and contained breathing apparatus) on task completion time and task accuracy, have been identified in previous studies on Level A protective chemical gear (Sargent & Murray, 2009) and included within IMPRINT. Task selection strategies (in addition to STOM) and overload thresholds have been included in IMPRINT (Little et al., 1993).

## S-PRINT BETA TEST

### Purpose

To evaluate the usability and effectiveness of the prototype tool, the training program, and model predictions, we conducted a beta test at NASA Johnson Space Center. Ten participants, representative of potential users, volunteered to attend training, perform tasks, and provide feedback on the tool. These participants included representatives from Behavioral Health and Performance, Space Human Factors, Flight Controllers, and Biomedical Engineers. The beta test version included the IMPRINT model of an operator using a fire detection / locator / suppression system.

### Methods

We conducted a 3-hour training session, in which we explained the purpose of the project, demonstrated all features of the tool, and showed how to create and run scenarios, and evaluate results. All participants received handouts of the presentations. They each had a laptop computer with the beta version of S-PRINT, so they could follow along using the tool.

Five tasks were assigned during the training. These addressed different aspects of operator performance that can be evaluated with S-PRINT: fatigue conditions, automation modes, automation failure types, task factors, and an “across the waterfront” test of different scenarios (best case to worst case). All participants volunteered for 1 task, so 2 participants were assigned to each of the 5 tasks. Finally, all participants were given a questionnaire regarding the overall performance and ease use of the tool, and understandability of results.

### Results

Of the 10 participants, 8 provided feedback on the questionnaire and created the data files of results for their tasks (2 were unable to continue participating). We obtained results for all of the 5 tasks. Key findings were that result files needed to be improved, the fatigue model (for 1 participant) produced identical results across vastly different conditions, and the instructions and handouts were necessary for participants to complete their tasks.

### Updates to the Tool

Based on the findings of the beta test, several changes were made to the tool. First, the reporting capability was

improved. We provided a button so users could access reports directly from S-PRINT, and we provided a reduced set of results deemed most useful to S-PRINT users, based on their feedback. We grouped the reports into four categories (input, overall performance, fatigue effects, and task selection) on the report selection window to help users predict the content. Further, we pruned the data provided in the reports to include only those factors stated to be of relevance by S-PRINT users.

With the Fatigue model, we realized that two pull-down menus were used for sleep histories. One was used to *create* a custom sleep history, and one was to *select* the sleep history. Users could easily choose the wrong menu. Our solution was to make a unique tab to *create* sleep histories, and another tab to *assign operator factors*, including individual sleep history.

The finding regarding the importance of detailed instructions led us to include a help system within S-PRINT. The help system describes the capabilities of the tool, demonstrates how to perform specific tasks, and shows how to compare the results of different model runs.

## Summary

The S-PRINT tool was found to be useful in terms of the questions it helped researchers address and the results it provided. Using IMPRINT as the basis for the tool allowed us to leverage sophisticated workload models, existing fatigue models, performance shaping factors, and reporting capabilities. Our work extended these models by accounting for fatigue effects on complex task performance, sleep inertia, and providing a new strategy for selecting tasks in an overload condition. In addition, we implemented a new capability to assess HAI. The IMPRINT human performance modeling environment provides end users with a product that can be used by both novices and modeling experts.

## MODEL VALIDATION

One additional, and critical, step in this research effort was an empirical validation study. This was performed to gather actual human performance data in the same conditions evaluated by the human performance model. The validation effort, described in Wickens, Vieane, et al., 2015, replicates most elements of the scenario described above. Participants worked with different modes of operation using a robotic arm simulation and a process control simulation. In the experiment, the process control system unexpectedly failed, thus creating a workload transition.

In the validation effort, we compared model predictions of operator performance with actual empirical data. These comparisons included task completion times or robotic arm trajectories completed, errors made, and operator workload. In particular, the task switching model was evaluated in more detail. Correlations were used to determine the degree to which the model correctly predicts differences in performance across the conditions. The validation study provided us an opportunity to improve the model so it delivers more accurate predictions of operator performance.

## CONCLUSIONS

In conclusion, human performance modeling provides a powerful tool for predicting operator and system performance in unknown situations. By basing the models on empirical operator performance, and conducting validation studies, we provide evidence that these models are effective prediction tools. In addition, to be beneficial to researchers, it is necessary that these models be easy to use. This paper describes an approach to developing an easy-to-use model-based tool, and to developing and empirically validating research-based models.

## ACKNOWLEDGMENTS

This work was supported by NASA NNX12AE69G. Brian Gore, Sandra Whitmire, and Jessica Marquez were the technical monitors. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of NASA.

## REFERENCES

- Allender, L., (2000). Modeling Human Performance. In M. Chinni (Ed.) Proc of the Military, Govt and Aerospace Simulation Symp, 2000 Advanced Simulation Tech Conf (Wash, DC) 139-144.
- Gutzwiller, R., Wickens, C.D. & Clegg, B. (2014). Workload Overload Modeling. *HFES Proceedings*.
- Hursh, S.R. (2003). *Sleep, Activity, Fatigue and Task Effectiveness (SAFTE) Model and the Fatigue Avoidance Scheduling Tool (FAST)*. ARFL: Sci and Tech for Tomorrow's Aero Force, Warfighter Fatigue Countermeasures.
- Little, R., Dahl, S., Plott, B., Wickens, C., Powers, J., Tillman, B., Davilla, D., Hutchins, C. (1993). Crew Reduction in Armored Vehicles Ergonomic Study (CRAVES) Final Report, Alliant Techsystems, Inc., Hopkins, MN, Tech. Rep. ARL-CR-80.
- Onnasch, L., Wickens, C., Li, H. & Manzey, D. (2013). Human Performance Consequences of Stages and Levels of Automation. *Human Factors*.
- Sargent, R.A. & Murray, S.L. (2009). Consequence Management Enhancements to Human Systems Integration (HSI) Analytic Tools. Final Report for the Leonard Wood Institute, Sub award number LWI 28-1217.
- Sebok, A., Wickens, C., Clegg, B. & Sargent, R. (2014). Using Empirical Research and Computational Modeling to Predict Operator Response to Unexpected Events. *HFES Proceedings*.
- Whitmire, A., Leveton, L., Barger, L., et al. (2009). Risk of performance errors due to sleep loss, circadian desynchronization, fatigue, and work overload. *Human Research Program Requirements Document*, HRP-47052, Rev. C.
- Wickens, C., Clegg, B., Vieane, A. & Sebok, A. (2015). Complacency and Automation Bias in the Use of Imperfect Automation. *Human Factors*.
- Wickens, C., Gutzwiller, R., & Santamaria, A. (2015) Discrete task switching in overload: A meta-analyses and a model. *International Journal of Human Computer Studies*.
- Wickens, C.D., Hooley, B.L., Gore, B.F., Sebok, A., & Koenecke, C. (2009). Identifying Black Swans in NextGen. *Human Factors*. 51: 638-651
- Wickens, C.D., Hutchins, S.D., Laux, L. & Sebok, A. (2015). The Impact of Sleep Disruption on Complex Cognitive Tasks. *Human Factors*.
- Wickens, C.D., Laux, L., Hutchins, S.D. & Sebok, A. (2014). Effects of sleep restriction, sleep inertia, and overload on complex cognitive performance before and after workload transition. *HFES Proceedings*.
- Wickens, C.D. & Sebok, A. (2014). Flight deck models of workload and multitasking: An overview of validation. Chpt 5 in M.A. Vidulich, P.S. Tsang & J.M. Flach (Eds.) *Advances in Aviation Psychology, Volume 1*. Burlington, VT: Ashgate Publishing Company. 69-84.
- Wickens, C., Vieane, A., Clegg, B., Sebok, A., & Janes, J. (2015). Visual attention allocation between robotic arm and environmental process control. *HFES Proceedings*.