Methods for Including Human Variability in System Performance Models

Randall J. Hodkin Jr. & Dr. Michael E. Miller
Air Force Institute of Technology, Department of Systems Engineering and Management

Abstract

To understand system performance, it is rational to consider all system components, including the humans involved in the control or maintenance of the system. Previous research has included human performance by modeling human tasks as events within Discrete Event Simulation (DES) models. These models typically represent the variability of task performance times and error rates by calculating the mean and variance across multiple individuals. Such approaches assume independence of task performance measures between individuals, but evidence exists which indicates that task performance measures are correlated between individuals. The current research seeks to understand methods to account for performance variability within DES models. A taxonomy of potential methods to address variability in DES models is developed and discussed. Among the findings derived through development of this taxonomy is the need to differentiate models of performance envelopes from models of average system performance and alternatives for modeling the human when predicting each class of performance.

Keywords: Discrete Event Simulation, Variability, Performance Envelopes, Mean Performance

Introduction

In its 2016 report on modeling, Sandia National Laboratories concluded that Model Based Systems Engineering (MBSE) provides significant advantage to project performance (Carroll & Malins, 2016). However, to obtain useful information from a simulation, modelers must adopt an approach appropriate for the specific engineering problem and capture sufficient system details to address the problem. Since humans interact with most systems, it is often important to include representations of human attributes or behaviors within system models when attempting to estimate system performance. In fact, one study by Baines et al in 2004 found that simulation results including Human Performance Models (HPM) could vary by as much as 35% when compared to results where human factors were not considered (Baines, Mason, Siebers, & Ladbrook, 2004).

Discrete Event Simulation (DES) environments, such as the Improved Performance Research Integration Tool (IMPRINT), have been used to predict human and system performance (Mitchell, 2000). These models typically require the modeler to decompose mission segments into a network of interdependent, discrete human activities or tasks and system activities. Generally, each human task and system activity is modeled as a discrete event having a time distribution and a probability of successful task completion. While several studies have illustrated that DES models can accurately predict mean times for mission completion, the models often underestimate variability in mission completion times (Kim, Miller, Rusnock, & Elshaw, 2018; Goodman, Miller, Rusnock, & Bindewald, 2017). A potential reason for this deficiency is the practice of collapsing task times across multiple individuals to obtain a mean while also assuming that each task is independent of the time required to complete other tasks.

In anthropometrics, it has long been recognized that some human measures are related to others with varying degrees of correlation. Initially, systems considering physical measures were designed to accommodate individuals with average anthropometric dimensions, assuming this strategy would produce systems that adequately accommodate most individuals. However, in 1952 Daniels showed that human anthropometric measures could have varying degrees of correlation with each other (Daniels, 1952). Specifically, Daniels demonstrated that from 4,000 pilots, not even one was within 30 percent of the mean in every one of ten physical dimensions, indicating that there was no "average" pilot. Since this study, anthropometric modeling has primarily shifted to defining a range or envelope of dimensions that accommodate a desired range of humans. Several statistical modeling methods that account for correlated measures were developed to define this envelope of human dimensions (Brolin, 2016; Kuo, Wang, & Lu, 2020; Hsiao, 2013; Kim & Whang, 1997; Jung, Choi, Lee, You, & Kwon, 2021; Zehner, Meindl, & Hudson, 1989).

Similarly, human performance varies across numerous dimensions, with some humans performing certain tasks consistently faster or more accurately than others. The effect is that, as with anthropometric dimensions, an individual's performance measures of similar tasks can be correlated. Additionally, comparably skilled people may experience correlation in task performance across individuals. Therefore, task performance is subject to correlation both within and across individuals. Since present human modeling is limited in these situations, this research explores methods for including task performance correlation and variability in DES models. Specifically, the current research undertakes a literature review to develop a taxonomy of approaches for quantifying or including human variability in DES models.

Background

DES as a human modeling approach simulates time-dependent real-world processes as events, updating system parameters during the events and approximating the system as fixed between events (Mathworks, 2021). It is particularly well suited for problems in the HPM domain since human interaction with a system can be characterized by procedures and broken down into task networks (Laughery, 1999). Each task in the network represents a time-dependent discrete event where human performance parameters are calculated. There are many methods for calculating system parameters during a discrete event and updates can either be programmed as deterministic or stochastic (Alion Science and Technology, 2018). Since every human has unique physical and cognitive limitations which can fluctuate over time depending on environmental and individual stressors, human performance is subject to between-subject and within-subject variation that typically requires stochastic DES methods (Belyavin & Fowles-Winkler, 2003).

A common source of uncertainty in human modeling is the variability of performance between individuals, which is often captured in DES as probably distributions that are developed from theory or more commonly determined using empirical data (Batarseh, 2010). To model with stochastics, an engineer must develop a probability distribution and parameters through analysis of empirical data (Greasley, 2016). Unfortunately, it is common to assume variable independence when developing distributions since this assumption simplifies analysis and modeling (Kruskal, 1988). In fact, some of these simplifications are inherent in DES software and can contribute to a modeler assuming independence. To better understand how independence can be assumed so easily it is important to understand the process for applying DES to HPM.

Modeling Human Performance Using DES

There are many ways different ways to conceptualize the steps in creating a simulation model. However, these models often include the five phases described by Allen (Allen, 2011). The first three phases outline model development while the final two phases are dedicated to model application. These five phases include:

Phase 1: Define (who, what, when, how)
Phase 2: Input Analysis (data collection and fitting distributions)
Phase 3: Simulation/Calculation (create/validate prediction models)
Phase 4: Output Analysis (alternative comparisons)
Phase 5: Decision Support (charts, tables, reports)

The primary model components that must be defined during phase one are the tasks required to complete the procedure being modeled, events that change the state of the system, entities that perform or participate in tasks, and state variables that define what is happening within a system at a given point in time (Banks, 1998). Task network modeling is an approach to understand what is to be modeled in DES where individual performance decomposed into a sequence of tasks. This process requires that a modeler possess knowledge about the system. If system knowledge is limited, then a task analysis may be required to appropriately scope the model and define model components (Rusnock, 2021). Simulation entities must be further defined by attributes that characterize how they interact with tasks and events must also be further defined to include proper sequence that maintains the systems physical cause-and-effect constraints (Banks, 1998). The amount of task decomposition and the overall model scope depends on the engineering problem posed (Laughery, 1999). Phase one of WSC model development is complete when all the components are defined at a level sufficient to support the modeling scope.

Transitioning to phase two, data is collected to quantify the time distributions, logic, and other relationships. This data is then analyzed to define the time distributions. For human modeling, this usually includes gathering data for multiple individuals, plotting the data to characterize the distribution of the data, and fitting probability distributions to the data. This last step often includes the use of analysis tools such as Q-Q plots and tests, such as the Kolmogorov-Smirnoff (K-S) test, to determine the best distribution, as well as calculating the parameters for the probability distribution.

To complete the process of developing a DES HPM, phase three includes creation of the tangible model. The process is comprised of creating the task network, entities, state variables, and entering the task probability parameters or any other parametric relationships. For human modeling in tools such as the IMPRINT this can also include setting up performance shaping parameters (PSF) to adjust human performance for changes in human, system, or environmental states. Additionally, IMPRINT permits one to estimate the workload for each task as well as how it influences simulation performance, and the tool must be properly configured to provide these values where needed. Finally, once the model is fully created and all the appropriate details are entered, the last step in this process is to validate the HPM.

When working through the three phases of the WSC development process, an engineer is determining all the information that will be used to approximate human behaviors. It is important that all the input parameters be meaningfully derived from theory, or more likely, determined from empirical data. For this reason, input analysis is prone to mistakes that can affect the accuracy of predictions and care should be taken when performing analysis to determine input parameters. Incorrectly assuming independence, lack of knowledge about dependencies, missing data, and measurement error all increase model uncertainty and reduce the accuracy of predictions if not appropriately considered (Batarseh, 2010).

DES and The Assumption of Independence

Unfortunately, it is common practice to assume variable independence when developing probability distributions (Kruskal, 1988). Although event independence is not necessarily inherent in DES, this common assumption in simulations stems from four influences. First, DES assumes discrete events that are each isolated in time and defined by probability distributions which assume independent and identically distributed data (Biller & Gunes, 2010; Corlu, Akcay, & Xie, 2020). Second, there is often insufficient data to characterize dependencies between tasks (Batarseh, 2010; Corlu, Akcay, & Xie, 2020). Third, the assumption of independence simplifies model construction by reducing the work required to determine input parameters (Kruskal, 1988; Romeu, 2006). Finally, while many input processes may exhibit dependence, DES software has not been updated to include an inherent method to handle variable dependency and it can only be implemented through difficult coding or iterative simulations (Corlu, Akcay, & Xie, 2020). Thus, new methods to manage dependencies can be difficult to understand and implement in existing software (Biller & Ghosh, Multivariate Input Processes, 2006).

It is important to point out that variability in HPM differs from many other domains. Human modeling seeks to understand the performance of a system given distinct interactions with multiple individuals. Many people will perform tasks within a system multiple times under different physical and mental conditions, and therefore HPM often needs to capture both within-subject and between-subject variability. However, it is common that both within-subject and between-subject variability is modeled using a single distribution. But, since each human participates in multiple tasks, an individual's specific abilities or state often influence not only their performance on the current task but also affect their performance in other tasks in a similar manner.

Therefore, human task performance is not truly independent and the amount of dependence between ability or state and performance is not static but can also vary based on the type of task performed. In fact, research has been performed to decompose tasks into a taxonomy of relatively independent sensory, cognitive, psychomotor, or physical tasks where performance is highly correlated (Fleishman, 1967; Furnham, von Stumm, Makedrayogam, & Chamorro-Premuzic, 2009). Thus, an individual who generally performs well at psychomotor tasks compared to his or her peers will frequently perform well at similar tasks in relation to their peers. The implication is that a DES model of performance across tasks will be correlated based on an individual's ability and state. Since task times are generally summed for all events, the total time will depend on correlations that should be included in the simulation. Thus, it is clear the traditional approach of combining data to find a distribution for each task will produce a model which has the highest likelihood of predicting the average task time but will reduce predictions of the system variability. Therefore, it is important to account for the correlations between an individual's skill and performance as we discuss methods that can adjust HPM to better predict variability.

Methods

A series of literature searches was conducted to develop a taxonomy for modeling population variability in HPM DES. The goal of initial searches was to better understand the steps used to develop a DES model, which steps might be modified to increase the accuracy of variability predictions, and to understand the factors that contributed to the need for modified steps. A final literature search was performed, applying the knowledge from previous searches, to determine approaches that could be used to better represent population variability in HPM DES models. From this final literature review, a taxonomy was then constructed to represent DES modeling approaches which provide more accurate predictions of human variability. The taxonomy also provides a quick visual reference to a range of tools, of varying prediction accuracy and rigor to implement, that can be employed at a modeler's discretion based on project needs.

All literature reviews conducted were structured to provide academic publications about DES and variability between the start of heavy commercial DES use in 1970 and the current date. Literature searches were completed using Google Scholar, EBSCO Discovery Services, and Academic Search Ultimate. Reviews followed the steps of examining abstracts, determining significance, and full review of important papers. Other relevant papers were also identified through a review of references from papers discovered during the literature search process. Some of the papers discovered through reference review were outside the search period but were still relevant and therefore cited as references for this paper.

Searches were conducted using key words, including DES in combination with terms to include human modeling, human performance, modeling people, independence assumption, input modeling, variability, uncertainty, and modeling techniques. The final literature search focused on DES, human variability, within-subject, between-subject, and modeling variability. Searches for these key words directly returned publications cited in this paper or returned papers that led to others through a reference review.

Results

Evidence has shown that the casual assumption of independence and limited capabilities available in many existing DES software are not adequate for accurately predicting population variability (Batarseh, 2010; Jung, Choi, Lee, You, & Kwon, 2021). As systems become increasingly complex the interactions humans have with them are also becoming more intricate. Hardware and interfaces are being tailored to accommodate individual differences in capabilities by balancing workload between humans and non-human agents so overload can be avoided. The result is that the demand for accurate human variability predictions is increasing and new approaches for modeling variability are necessary to meet the need.

The articles reviewed during this research tended to fall into one of two categories that described methods for either quantifying or reducing uncertainty, or variability, in a model. As shown in Figure 1, the first classification is based on whether a method actively attempts to reduce model uncertainty or whether it seeks to quantify the amount of uncertainty expected in model results.

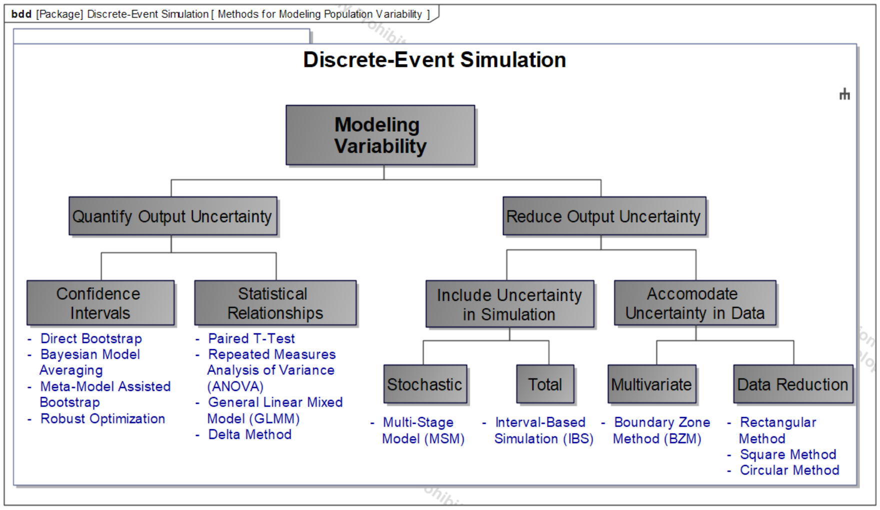

Figure 1. Taxonomy of Methods for Modeling Population Variability in DES

While methods for quantifying uncertainty have been around since the 1970's or earlier, approaches to reduce output uncertainty only started to emerge in the early 2000's with articles like the one published by Belyavin et al in 2003. This paper the discussion of ways to reduce uncertainty by introducing additional variability to models.

The two broad approaches of quantifying or reducing output uncertainty are quite different in specific technique and results but achieve the same overall goal of understanding variability in DES models. quantification methods return DES model outputs and report their associated confidence interval (CI) while reduction methods return outputs that have typically been adjusted to fall within a certain CI. From a practical standpoint, this means that methods for quantify uncertainty are good for risk assessment and decision support while uncertainty reduction methods are a better choice for engineering design where results will feed other analyses. As the approaches to quantifying output uncertainty are relatively well understood, it is useful to focus on methods to reduce output uncertainty.

The taxon for reducing output uncertainty is further decomposed into two categories. The first permits uncertainty to be included in a simulation directly while the second seeks to understand uncertainty due to between-subject differences. The next level of taxon further divides these techniques into four distinct categories with each technique providing a method for modeling and understanding data using simulation loops or iterative simulation to model within-subject and between-subject variability. However, only the multi-stage model (MSM) and Interval Based Simulation (IBS) methods for representing total uncertainty have currently been discussed in literature as a potential methods specific to DES.

Methods in the stochastic taxon apply multi-stage stochastic models to predict performance. For example, Belyavin et al suggested a multi-stage stochastic model where operator state is predicted and used to influence the normal DES model, thereby introducing additional variability through an extra modeling stage (Belyavin & Fowles-Winkler, 2003). Similarly, IBS introduces imprecise probabilities via iterative simulation to represent total uncertainty (reducible & epistemic) and better capture system variability.

Methods like MSM or IBS that introduce uncertainty terms to a model still relate to prediction of the total variability about the mean, but another approach is to attempt to understand the envelope of performance. Techniques that accommodate uncertainty have previously been applied in the anthropometric domain and permit one to understand representative individuals whose performance encompass a defined envelope of performance. The task performance of these representative individuals can then be modeled using appropriate probability distributions, and their performance distributions iteratively modeled within a DES. Through this approach, one can understand the boundary of performance across a population of individuals without modeling the effect of each individual within a population.

Discussion and Conclusions

Overall, this research did not identify any current methods in DES literature which address the lack of dependence in task performance across individuals. Belyavin's MSM approach suggests methods to improve within-subjects variability and approaches such as IBS might be used to address this issue, but its implementation would require substantial modification of existing DES tools to provide a practical approach. However, one can envision a combination of MSM and anthropometric tools like BZM to select representative humans for simulations where an inner loop simulates the performance of an individual and an outer loop simulates individual differences to capture a better estimate of variability.

As part of this research, it appears some classes of HPM may benefit from performance envelopes rather than average system performance. As with anthropometrics, there are classes of HPM for which the goal may be to accommodate a range of performance rather than predict the average result. Methods like BZM that were originally designed for multivariate anthropometric accommodation can be leveraged to achieve this result. BZM and similar approaches may not only help with variability predictions but might also open the door to new DES approaches that can predict human task performance envelopes.

Overall, HPM tools like DES have been updated with new interfaces and some features like PSF and workload calculations, but it may be time to consider more drastic changes to improve prediction accuracy. Methods like IBS and BZM documented in this research offer ways that DES can be modernized to allow consideration of human performance envelopes when the situation dictates.

References

Alion Science and Technology. (2018). *IMPRINT Pro User Guide.* Louisville, CO: Alion Science and Technology, HSI DIvision.

Allen, T. (2011). *Introduction to Discrete Event Simulation and Agent-based Modeling.* London: Springer-Verlag.

Baines, T., Mason, S., Siebers, P., & Ladbrook, J. (2004). Humans: the missing link in manufacturing. *Simulation Modelling Practice and Theory*, 515-526.

Banks, J. (1998). *HANDBOOK OF SIMULATION Principles, Methodology, Advances, Applications, and Practice.* New York: John Wiley and Sons, INC.

Batarseh, O. (2010). *An Interval Based Approach to Model Input Uncertainty in Discrete-Event Simulation.* Orlando: University of Central Florida.

Belyavin, A., & Fowles-Winkler, A. (2003). Subject Variability and the Effect of Stress in Discrete-Event Simulation. *Proceedings 15th European Simulation Symposium.*

Biller, B., & Ghosh, S. (2006). Multivariate Input Processes. *Elsevier Handbook in OR & MS, Vol 13*, 123-153.

Biller, B., & Gunes, C. (2010). Introduction to Simulation Input Modeling. *2010 Winter Simulation Conference*, (pp. 49-58).

Brolin, E. (2016). *Consideration of Anthropometric Diversity - Methods for Virtual Product and Production Development.*

Carroll, E. R., & Malins, R. J. (2016). *Systematic Literature Review: How is ModelBased Systems Engineering Justified?* Albuquerque, NM: Sandia National Laboratories.

Colombi, J., Miller, M., Bohren, J., & Howard, J. (2014). Conceptual Design Using Executable Architectures for a Manned Mission to Mars. *IEEE Systems Journal 2014*, 1-18.

Colombi, J., Miller, M., Schneider, M., McGrogan, J., Long, D., & Plaga, J. (2011). Model Based Systems Enegineering with Department of Defence Architectural Framework. *Systems Engineering 2011:4(3)*, 305-326.

Corlu, C., Akcay, A., & Xie, W. (2020). Stochastic simulation under input uncertainty: A Review. *Elsevier Operations Research Perspectives*, 1-16.

Daniels, G. (1952). *The Average Man?* Dayton: Wright-Patterson Air Force Base.

Fleishman, E. (1967). Development of a Behavior Taxonomy for Describing Human Tasks: A Correlational-Experiment Approach. *Journal of Applied Psychology*, 1-10.

Furnham, A., von Stumm, S., Makedrayogam, A., & Chamorro-Premuzic, T. (2009). A Taxonomy of Self-Estimated Human Performance. *Journal of Individual Differences*, 188-193.

Goodman, T., Miller, M., Rusnock, C., & Bindewald, J. (2017). Effects of Agent Timing on the Human-Agent Team. *Cognitive Systems Research 2017:46*, 40-51.

Greasley, A. (2016). Methods of Modelling People using Discrete-event Simulation. *6th International Conference on Simulation and Modeling Methodologies, Technologies and Applications* (pp. 312-317). SCITEPRESS – Science and Technology Publications.

Hsiao, H. (2013). Anthropometric Procedures for Protective Equipment Sizing and Design. *Human Factors 2013:55(1)*, 6-35.

Jung, K., Choi, Y., Lee, B., You, H., & Kwon, O. (2021). A Boundary Zone Method for the Generation of Multivariate Representative Humanoids. *MDPI Applied Sciences*.

Kim, J., & Whang, M. (1997). Technical Note: Development of a set of Korean Manikins. *Apllied Ergonomics 1997:28*, 407-410.

Kim, S., Miller, M., Rusnock, C., & Elshaw, J. (2018). Spatialized audio improves call sign recognition during multi-aircraft. *Elsevier Applied Ergonomics*, 51-58.

Kruskal, W. (1988). Miracles and Statistics: The Casual Assumption of Independence. *Journal of the American Statistical Association*, 929-940.

Kuo, C., Wang, M., & Lu, J. (2020). Developing Sizing Systems using 3D Scanning Head Anthropometric Data. *Measurement Journal of the International Measurement Confederation*.

Laughery, R. (1999). Using Discrete-Event Simulation to Model Human Performance in Complex Systems. *Winter Simulation Conference*, (pp. 815-820).

Madni, A., & Madni, C. (2018). Architectural Framework for Exploring Adaptive Human-Machine Teaming Options in Simulated Dynamic Environments. *Systems 2018:6(4)*, 44.

Mathworks. (2021, December 8). *Understanding Discrete Event Simulation*. Retrieved from Mathworks: https://www.mathworks.com/videos/understanding-discrete-event-simulation-part-1-what-is-discrete-event-simulation--1494873178760.html

McManus, H., Richards, M., Ross, A., & Hastings, D. (2007). A Framework for Incorporating "ilities" in Tradespace Studies. *A Collection of Technical Papers: AIAA Space 2007 Conference* (pp. 941-954). AIAA.

Mitchell, D. (2000). *Mental Workload and ARL Workload Modeling Tools.* ARL.

Romeu, J. (2006). Teaching Engineering Statistics to Practicing Engineers. *7th International Conference on Teaching Statistics* (pp. 1-6). Salvador, Bahia, Brazil: ICOTS-7.

Rusnock, C. (2021). Human Systems Modeling: Task Analysis. Dayton, OH, USA.

Watson, M., Rusnock, C., Miller, M., & Colombi, J. (2017). Informing System Design Using Human Performance Modeling. *Systems Engineering 2017:20(2)*, 173-187.

Zehner, G., Meindl, R., & Hudson, J. (1989). *A Multivariate Anthropometric Method for Crew Station Design.* DTIC.