INL/CON-16-38029
PREPRINT

SIMULATED HUMAN ERROR PROBABILITY AND ITS APPLICATION TO DYNAMIC HUMAN FAILURE EVENTS

13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13) (2016)

Sarah M. Herberger and Ronald L. Boring

October 2016

SIMULATED HUMAN ERROR PROBABILITY AND ITS APPLICATION TO DYNAMIC HUMAN FAILURE EVENTS

Sarah M. Herberger[1] and Ronald L. Boring[1]

[1] *Idaho National Laboratory: P.O. Box 1625, Idaho Falls, Idaho 83415-3818,* sarah.herberger@inl.gov

Human reliability analysis (HRA) methods typically analyze human failure events (HFEs) at the overall task level. For dynamic HRA, it is important to model human activities at the subtask level. There exists a disconnect between the dynamic subtask and static task levels that presents issues when modeling dynamic scenarios. For example, the SPAR-H method is typically used to calculate the human error probability (HEP) at the task level. Quantification in SPAR-H does not necessarily translate to the subtask level. In this paper, two different discrete distributions were considered for each of the eight SPAR-H performance shaping factors (PSFs) to define the frequency of each PSF level. The first distribution considered was a uniform discrete distribution that presumed the frequency of each PSF level was equally likely. The second non-continuous distribution took the frequency of each PSF level as identified from a subjective assessment of the HERA database. These two different approaches were created, so that the HEP could be calculated and a distribution identified. The HEP distribution that appears closer to the previously observed HEP, a log-normal centered on 1E-3, is the more desirable. Each HEP distribution then has median, average, and maximum HFE calculations applied. To calculate these three generic human actions—HFE A, B and C—are generated from the PSF level frequencies comprised of subtasks. The summary statistics for the HFE are applied as aggregate functions at each PSF level and then the HEP is calculated. The same data set of subtask HEPs yields starkly different HEPs when aggregated to the HFE level in SPAR-H. Assuming that each PSF level in each HFE is equally likely creates an unrealistic distribution of the HEP that is centered at 1. Next the observed frequency of PSF levels was applied with the resulting HEP behaving log-normally with a vast majority of the values under 2.5% HEP. The median, average and maximum HFE calculations did yield different answers for the HFE. The HFE maximum grossly overestimates the HFE, while the HFE distribution occurs less than HFE median, and greater than HFE average.

I. INTRODUCTION

The legacy of human reliability analysis (HRA) is that almost all methods to date have been static (Ref 2), meaning the approaches model a given set of human failure events (HFEs) but do not adapt to changing conditions in the model. Just as the adaptation from design-basis to beyond-design basis is difficult for static methods, the problem is made more complex when introducing dynamic HRA methods, which look at the emergent evolution of an event instead of analyzing a prescripted set of scenarios. The promise of dynamic methods is that they will be able to model performance more completely than the expert judgment processes required for completing static HRAs. The downside of dynamic methods is the increased methodological and implementational complexity that leads to longer calculation times. The general challenge of making HRA dynamic is increased multifold when dynamic methods must tackle the inherent uncertainty of severe accidents. Not only is the method complexity increased, but so is the modeling complexity.

Static methods are based on analyzing human performance for a pre-defined set of tasks that are generally clustered as HFEs. The challenge in extrapolating from these HFE snapshots to dynamic models is that many of the basic assumptions of these methods have not been validated for dynamic applications. For example, as depicted hypothetically in Fig 1, a sequence of events can be parsed in many ways. The horizontal axis divides the event along a chronological progression, in this case in terms of minutes. The dotted vertical lines demark subtasks during the sequence of events. Finally, the blue boxes denote HFEs. Each minute reveals a different outcome in terms of the dynamic HEP calculation. Similarly, the subtasks and HFEs track the changing HEP. Yet, HRA methods are not designed to track at all three levels of delineation. An HRA method that is applied successfully to three sequential HFEs as part of an event progression may not adequately cover further delimiting the HFE into 9 subtasks or 10 minute-long time slices. To model the event progression, however, it is necessary to model the HFE at a finer granularity corresponding to the 9 subtasks or 10 time slices. The static HRA method may not lend itself to these different units of analysis. Moreover, the error quantification approach used may not prove accurate for the different unit of analysis.

To frame the event progression in Fig 1 differently, consider the case of a major flooding incident. Major damage to the plant is sustained around the 4-minute mark along the timeline. HFE_1 corresponds to the pre-initiator, HFE_2 encompasses the initiating event, and HFE_3 spans the post-initiator recovery. As can be seen, the human error probability (HEP) remains low during the pre-initiator period, surges during the initiating event, and remains high during the recovery period. Static HRA methods, which would tend to analyze the event in terms of the three HFEs, may not fully model the changes to operator performance within each HFE. For example, a sudden increase in stress that causes a surge in error during HFE_2 actually consists of three different slopes of the error plot—an initial relatively flat period, a rapidly rising period, and a plateau that shows signs of gradually declining. The flooding has differing effects on the plant and the operators, but conventional static parsing of the event may not fully map the dynamic progression of the event and the equally dynamic error curve associated with different tasks and time slices.

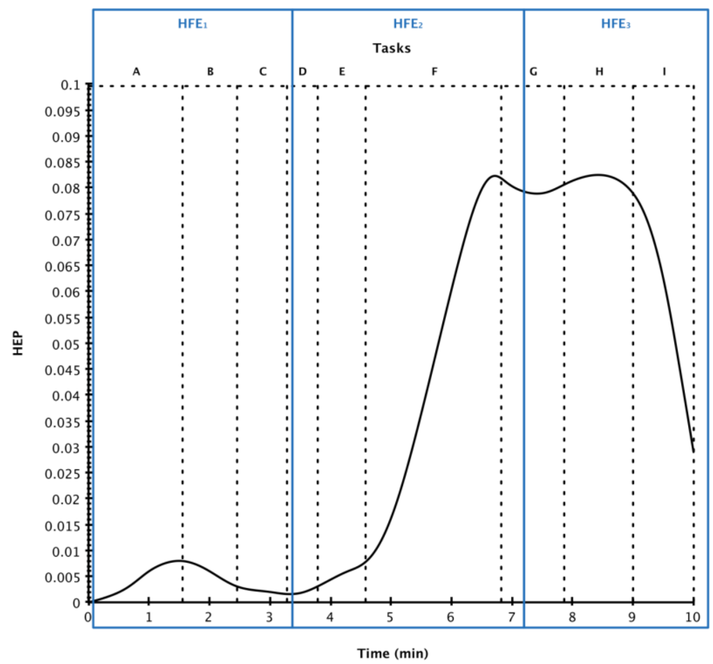

Fig 1. Human event progression according to time slices, subtasks, and HFEs.

This paper reviews what happens to HRA when the unit of analysis is changed from an HFE to a unit of analysis suitable for dynamic modeling. Underlying this discussion is the key assumption that dynamic HRA requires a finer grain of modeling precision than the HFE. Ideally, the HFE represents a thorough human factors subtask analysis (Ref 5; 7, and 8). The human reliability analyst will then quantify the event at the appropriate level of aggregation. HRA methods treat the unit of quantification differently. For example, the original HRA method, the Technique for Human Error Prediction (THERP, Ref 10) quantifies at the subtask level. In contrast, the Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) method (Ref 6) analyzes events at the HFE level, despite being derived from THERP (Ref 1). Ideally, the quantification approach should transfer between different framings of the event space. Additionally, associated with each HEP is also a measure of uncertainty. The uncertainty discussion centers on statistical considerations associated with propagating uncertainty over a large number of units of analysis.

II. SPAR-H FRAMEWORK

SPAR-H is a widely accepted method to determine the HEP based on expert estimation using calculation worksheets. Estimations are carried out using weighted performance shaping factors (PSFs) and a standard diagnosis failure probability. In many HRA methods, including SPAR-H, context-specific probabilities are generated by multiplying a nominal HEP by multipliers representing the effect of specific context elements which were deemed relevant to the problem by the method developers. This has resulted in the following equation:

$$HEP = NHEP * PSF \tag{1}$$

where *HEP* is the human error probability for the HFE; *NHEP* is the nominal human error probability, which is assumed to be 1E-3 based upon the Action worksheet in SPAR-H; and *PSF* is the product of all eight PSFs in the method (Ref 6). PSFs come in many flavors, with SPAR-H defining: available time, stress and stressors, complexity, experience and training, procedures, ergonomics and human-machine interface, fitness for duty, and work processes. Each PSF has different levels with a corresponding multiplier for diagnosis and action as seen in Table 1.

Table 1. The SPAR-H PSFs with their respective levels, action multiplier, diagnosis multiplier, action frequency, uniform frequency, Human Event Repository and Analysis (HERA) frequency (Ref 4), uniform probability, and HERA action probability. P(F)=1 stands for the probability of failure is equal to 1.

PSF	PSF Levels	Diagnosis Multiplier	Action Multiplier	Uniform Action Frequency	HERA Action Frequency	HERA Action Probability	Uniform Action Probability
Available Time	Inadequate Time	P(F)=1	P(F)=1	91	5	0.009	0.167
	Available Time = Time Required	10	10	91	26	0.048	0.167
	Nominal Time	1	1	91	500	0.914	0.167
	Time Available > 5x the Time Required	0.1	0.1	91	10	0.018	0.167
	Time Available > 50x the Time Required	0.01	0.01	91	4	0.007	0.167
	Insufficient Information	1	1	91	2	0.004	0.167
Stress	Extreme	5	5	149	2	0.003	0.25
	High	2	2	149	92	0.154	0.25
	Nominal Time	1	1	149	500	0.839	0.25
	Insufficient Information	1	1	149	2	0.003	0.25
Complexity	Highly Complex	5	5	134	3	0.006	0.25
	Moderately Complex	2	2	134	31	0.058	0.25
	Nominal	1	1	134	500	0.933	0.25
	Obvious diagnosis	0.1	-	-	-	-	-
	Insufficient Information	1	1	134	2	0.004	0.25
Experience	Low	10	3	140	50	0.089	0.25
	Nominal	1	1	140	500	0.893	0.25
	High	0.5	0.5	140	8	0.014	0.25
	Insufficient Information	1	1	140	2	0.004	0.25
Procedures	Not Available	50	50	112	1	0.002	0.2
	Incomplete	20	20	112	20	0.036	0.2
	Available but Poor	5	5	112	40	0.071	0.2
	Nominal	1	1	112	500	0.891	0.2
	Diagnostic/symptom oriented	0.5	-	-	-	-	-
	Insufficient Information	1	1	112	0	0	0.2
Ergonomics	Missing / Misleading	50	50	107	3	0.006	0.2
	Poor	10	10	107	30	0.056	0.2
	Nominal	1	1	107	500	0.938	0.2
	Good	0.5	0.5	107	0	0	0.2
	Insufficient Information	1	1	107	0	0	0.2
Fitness for Duty	Unfit	P(F)=1	P(F)=1	127	0	0	0.25
	Degraded Fitness	5	5	127	8	0.016	0.25
	Nominal	1	1	127	500	0.984	0.25
	Insufficient Information	1	1	127	0	0	0.25
Work Process	Poor	2	5	160	120	0.188	0.25
	Nominal	1	1	160	500	0.782	0.25
	Good	0.8	0.5	160	19	0.030	0.25
	Insufficient Information	1	1	160	0	0	0.25

As per SPAR-H, Ref 6, *HEP* is calculated with the action or diagnosis multiplier value is substituted in for the respective PSF levels to produce the following equation:

$$HEP = NHEP * available\ time * stress * complexity * experience * procedures * ergonomics * fitness\ for\ duty * work\ process \quad (2)$$

where each PSF is substituted with the respective PSF level's multiplier. Of course, each level of a PSF is not equally likely. As such, the frequency of PSF level assignments was taken from Ref 4. Additionally, for the purposes of this exploratory analysis, only the SPAR-H Action worksheet PSF multipliers are used. The adjustment factor is applied when three or more PSFs are negative, as per equation (3):

$$HEP = (NHEP * PSF) / [NHEP * (PSF - 1) + 1] \quad (3)$$

A negative PSF is when a multiplier is larger than 1, and contributes to increasing the HEP. The probabilities and frequencies used in the differing simulation and analysis can be seen in Table 1.

III. HUMAN FAILURE EVENT SIMULATION

The HFE simulation is based on the probabilities of a PSF level in Table 1 and Equation (2). A simulation of 5,000 data points was run to represent the distribution of a single task. This is then repeated for Tasks A, B, C, D, E, and F so that there are a total of 30,000 data points in Fig 2. Tasks A, B, C, D, E, and F are seen as generic human actions that should be comparable to one another other than their differing PSF frequencies. Tasks A, B, and C come from a uniform PSF frequency and D, E and F come from the HERA frequencies. The frequencies used are from their respective Table 1 columns.

When the uniform PSF levels are implemented, distributions as in Fig 2 right are generated, and Tasks A, B and C visually appear to be similar and be strongly skewed toward an HEP of 1. To verify the similarity of the generic human tasks from the simulation, a one-way analysis of variance could be used to compare means of three or more groups. However, the distributions of the Tasks and HFE are clearly not normally distributed, thus a non-parametric approach, Kruskal-Wallis H test (KWH), is suggested for comparison purposes. When generic human Tasks A, B, and C are compared using a KWH with 2 degrees of freedom, a *p*-value of 0.8813 is received. Likewise generic human Tasks D, E, and F are also similar to one another; however, they are skewed toward an HEP of 0. These tasks are compared to one another using a KWH with 2 degrees of freedom the resulting *p*-value is 0.4027. Both of the *p*-values are severely not significantly different as they are greater than 0.05. They are considered generic human tasks and should be very similar to one another. Violin plots which display the distributions of Tasks A, B, C, D, E and F can be seen in Fig 2.

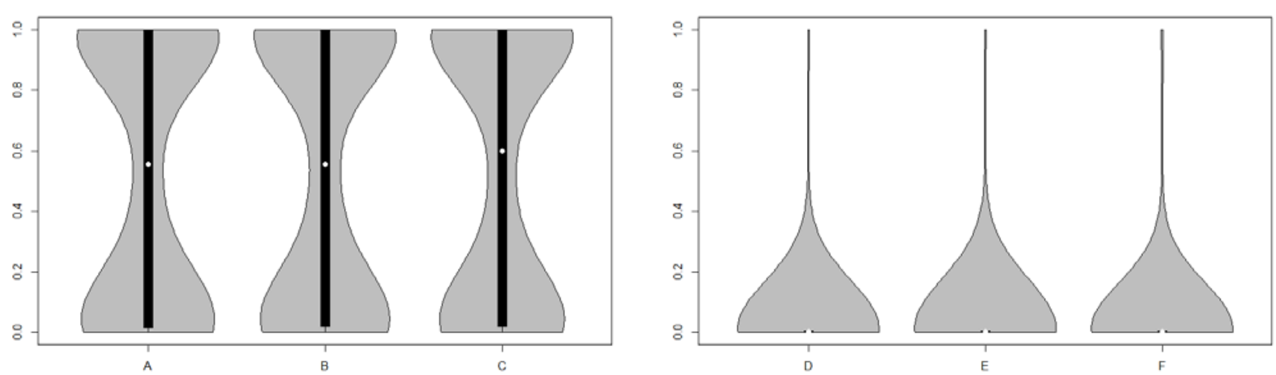

Fig 2. Violin plots of Tasks A, B and C assuming each PSF level is equally likely. Violin plots of Tasks D, E and F take into consideration PSF frequencies from Ref 4. Tasks A, B, C, D, E, and F are considered generic human actions and are simulated in the same manner, other than the PSF level frequencies. Each task was sampled 5,000 times from each PSF with frequencies.

Violin plots are displayed in Fig 2 and Fig 3, they are considered very useful for visualizing data because they are a boxplot with a histogram overlay. The boxplot identifies the quantiles, and the ends of the thick black bar in the middle of each violin plot of Fig 2 identify the 25th and 75th percentile. The thin black line is the whiskers in the boxplot going from the 25th to the minimum and the 75th to the maximum. Lastly the white dot in the thick black part of the box plot symbolizes the median. The images in Fig 2 and Fig 3 do not display an exemplary violin plot, because the data is so severely skewed. All analysis and graphical output were generated from R 2.2.3 (Ref 9).

Additionally, some anomalies occur in the data. Such anomalies occur when a PSF multiplier is P(F)=1 and when a HEP is greater than 1, even when the adjustment factor from equation (3) is use. A PSF multiplier of P(F)=1 transpires in two PSFs: available time and fitness for duty. A P(F)=1 automatically pushes the HEP to 1; however, the approach to the aggregate functions in this case needs special consideration. In order to quantify P(F)=1, equation (2) is solved for the respective PSF. If P(F)=1 occurs for both available time and fitness for duty it is assumed that they have equal bearing on the impending failure. This is necessary so that the aggregate functions can be empirically evaluated. An example of the quantification for the PSF multipliers is detailed in Table 2.

Table 2. Example of how the multipliers are quantified when P(F)=1 is present for available time and fitness for duty. Grey rows have P(F)=1 and the white rows have the multiplier values subbed in for the P(F)=1.

Available Time	Stress	Complexity	Experience / Training	Procedures	Ergonomics / human machine interface	Fitness for Duty	Work Process	HEP
P(F)=1	1	5	1	1	1	5	1	1
40	1	5	1	1	1	5	1	1
0.1	2	2	1	50	0.5	P(F)=1	1	1
0.1	2	2	1	50	0.5	100	1	1
P(F)=1	5	5	0.5	50	10	P(F)=1	5	1
0.016	5	5	0.5	50	10	0.016	5	1

Additionally there are combinations of PSF that can calculate a HEP greater than 1, when this occurs, the HEP is assumed to remain 1. The PSF multipliers remain at their original values and are not altered. An example of PSF multipliers producing an HEP greater than 1 is displayed in Table 3.

Table 3. Example of SPAR-H multipliers that the produced HEP is greater than 1.

Available Time	Stress	Complexity	Experience / Training	Procedures	Ergonomics / human machine interface	Fitness for Duty	Work Process	HEP
1	1	1	0.5	50	50	1	1	1
1	1	1	1	50	50	1	1	1

Specifically for the combination of PSF multipliers in Table 3 the adjustment factor from equation (3) is not applied because the number of negative PSFs is only 2. The HEPs would have been 1.25 and 2.5 respectively. However the HEP is assumed to be 1, as a human action cannot have a failure likelihood greater than 1.

Multiple tasks are often grouped as HFEs and SPAR-H assumes the unit of analysis is the HFE. If HFE_1 is comprised of Tasks A, B, and C (see Fig 1), there are then several ways to calculate the HFE based on a PSF multiplier or group of PSF multipliers. The maximum HFE calculation selects the largest PSF level values across three tasks. The assumption is that the analysis should capture the strongest manifestation of the PSF, even if the PSF changes across the evolution of the HFE. An example of this would be when a human reliability analyst decides to make a conservative or worst case estimation of the a changing set of tasks within a single HFE.

This HFE is then repeated with each respective aggregate function being applied at the PSF level across three tasks for: median, average, and multiplication. The methods are very similar to what intuition would produce when executed. The median takes the median PSF multiplier of three tasks. The average, the average of three tasks, and the multiplication approach takes the product of three tasks for a single PSF. An example of these aggregate function applied to three tasks for a single PSF, stress, is available in Table 4. The distributions for the different HFE aggregate functions can be seen in Fig 3.

Table 4. An example showing how aggregate functions are applied to the stress PSF of tasks A, B, and C. The same aggregate functions are used at the PSF level to quantify the HEP of tasks D, E, and F.

	Available Time	Stress	Complexity	Experience / Training	Procedures	Ergonomics / human machine interface	Fitness for Duty	Work Process	HEP	Max Stress	Median Stress	Average Stress	Multiplication Stress
Task A	1	5	1	3	5	50	1	1	0.79	5	5	3.6667	25
Task B	1	1	1	0.5	20	1	5	5	0.2				
Task C	1	5	2	0.5	1	1	40	5	1				

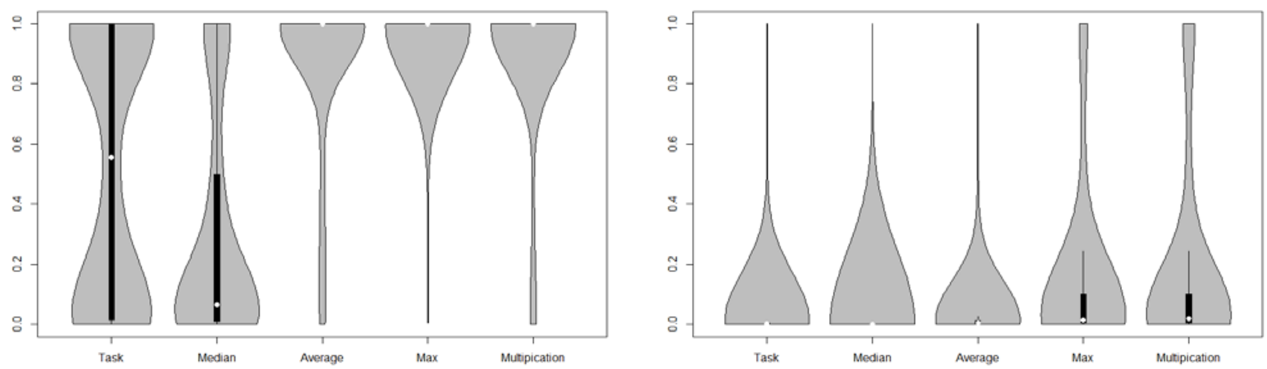

Fig 3. Violin plots of HFEs calculated in four different methods using the aggregate functions. (Left) Tasks generated from PSF levels that are equally likely, then the aggregate functions are applied. (Right) Tasks generated from PSF levels that are informed from the HERA data (Ref 4). The Maximum (max) calculation selects the largest of the three tasks. Median (med) selects the median value of the three tasks. Average (avg) calculates the average of the three tasks. The left is calculated using frequencies from Ref 4, while the right is calculated assuming a uniform frequency for all PSF levels.

Tasks A, B, C and their respective aggregate functions were compared using a KWH. This was then repeated for tasks D, E, and F and their associated aggregate functions. The comparisons, degrees of freedom, chi-square, and *p*-value for these analyses are displayed in Table 5.

Table 5. Results from the comparison using KWH.

Comparison	Degrees of Freedom (df)	chi-square	*p*-value
Task A, B, C, & Max	3	4862.2	< 0.001
Task A, B, C, & Median	3	137.12	< 0.001
Task A, B, C, & Average	3	3102.8	< 0.001
Task A, B, C, & Multiplication	3	3764	< 0.001
Task D, E, F, & Max	3	3950.4	< 0.001
Task D, E, F, & Median	3	1136.2	< 0.001
Task D, E, F, & Average	3	1387.3	< 0.001
Task D, E, F, & Multiplication	3	4415.5	< 0.001

Tasks A, B, C and Maximum HFE were compared using a KWH analysis and received ***p*-value < 0.001 with 3 degrees of freedom** (df). Tasks A, B, C and Average HFE were compared using a KWH and received ***p*-value < 0.001, df=3**. Both of these *p*-values indicate that Maximum HFE and Average HFE are significantly different from Tasks A, B, and C (Fig 4). Additionally, Tasks A, B, C and Median HFE were compared using a KWH and received ***p*-value < 0.001, df=3**. While still significant, visually and empirically Median HFE is the closest in distribution to the three tasks. The same results are found for Tasks D, E and F and their associated aggregate functions. Median PSF multipliers are the closest approximation to the task. A graphical representation can be seen in Fig 4. Generally, Maximum HFE overestimates Tasks A, B, and C and Average HFE underestimates Tasks A, B, and C. Again all 14 distributions, Task A, B, C, D, E, and F and their associated Max HFE, Median HFE, Average HFE and multiplication HFE can be seen in Fig 4.

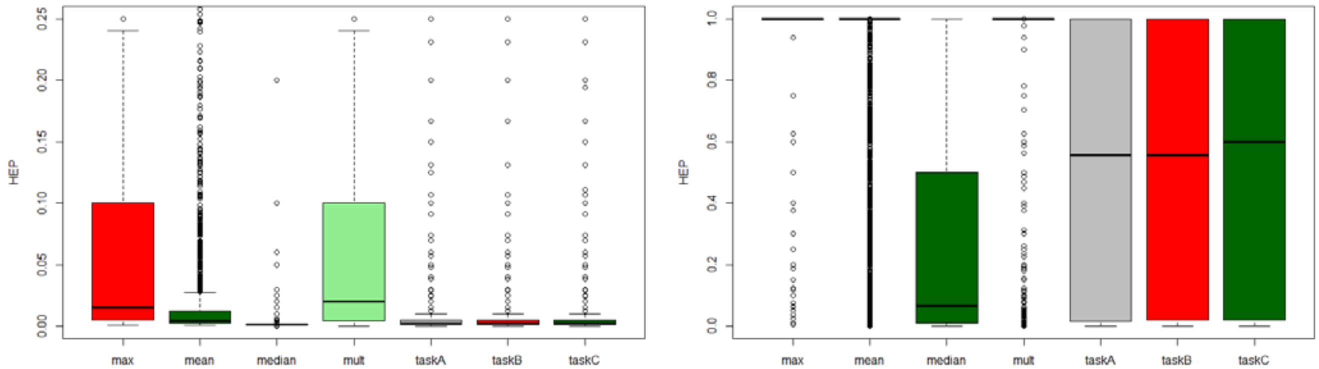

Fig 4. (Left) HFE Maximum, and HFE Average, HFE Median, HFE Multiplication, Tasks A, B, and C, with frequencies from a discrete uniform distribution. (Right) HFE Maximum, and HFE Average, HFE Median, HFE Multiplication, Tasks D, E, and F with frequencies from Ref 4. Please note the large difference in the y-axis range.

IV. CONCLUSION

This exploration into dynamic subtask to HFE task translation has provided examples of the process using the SPAR-H method. Dynamic task modeling is very difficult to peruse through the framework of SPAR-H; distributions associated with each PSF need to be defined, and may change depending upon the scenario. However it is very unlikely that each PSF level is equally likely as the resulting HEP distribution is strongly centered at 100%, which is unrealistic. Continuous distributions need to be identified for PSFs, to facilitate the transition to dynamic task modeling. Additionally discrete distributions need to be exchanged for continuous so that simulations for the dynamic HFE can further advance.

The SPAR-H decomposition shows approximation methods (median, average, max, multiplication) and applying SPAR-H to the sub-task level so a time series could be built. Based on these results SPAR-H quantification breaks down if the task level is not carefully controlled. Conceptually it is difficult to proceed with a dynamic model in SPAR-H given one of the PSFs is available time, rather than time impacting other relevant PSFs. The inaccuracy in SPAR-H HEP quantification only exists at the subtask modeling for dynamic HRA. The current level of analysis in existing HRAs using the method as defined would cause no need to worry about this issue nor need to revisit their quantification. It is expected that the concerns with ensuring the correct level of task decomposition for quantification apply across a wide variety of HRA methods beyond SPAR-H.

ACKNOWLEDGMENTS

Many thanks for input from Jeffery Einerson, Diego Mandelli, and the other staff at INL. Every effort has been made to ensure the accuracy of the findings and conclusions in this paper, and any errors reside solely with the authors. This work of authorship was prepared as an account of work sponsored by Idaho National Laboratory, an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately-owned rights. Idaho National Laboratory is a multi-program laboratory operated by Battelle Energy Alliance LLC, for the United States Department of Energy under Contract DE-AC07-05ID14517.

REFERENCES

1. Boring, R.L., Blackman, H.S. (2007). The origins of the SPAR-H method's performance shaping factor multipliers. Official Proceedings of the Joint 8th IEEE Conference on Human Factors and Power Plants and the 13th Annual Workshop on Human Performance/Root Cause/Trending/Operating Experience/Self Assessment, 177-184.
2. Boring, R., Mandelli, D., Joe, J., Smith, C., Groth, K. (2015). *A Research Roadmap for Computation-Based Human Reliability Analysis, INL/EXT-15-36051*. Idaho Falls: Idaho National Laboratory.

3. Boring, R., St. Germain, S., Banaseanu, G., Chatri, H., Akl, Y. (2015). Applicability of simplified human reliability analysis methods for severe accidents. *Proceedings of the 7th International Conference on Modelling and Simulation in Nuclear Science and Engineering.*
4. Boring, R.L., Whaley, A.M., Tran, T.Q., McCabe, P.H., Blackwood, L.G., Buell, R.F. (2006). *Guidance on Performance Shaping Factor Assignments in SPAR-H, INL/EXT-06-11959.* Idaho Falls: Idaho National Laboratory.
5. Electric Power Research Institute (EPRI). (1992). *SHARP1—A Revised Systematic Human Action Reliability Procedure, EPRI-101711*. Palo Alto: Electric Power Research Institute.
6. Gertman, D., Blackman, H., Marble, J., Byers, J., & Smith, C. (2005). *The SPAR-H Human Reliability Analysis Method, NUREG/CR-6883*. Washington, DC: U.S. Nuclear Regulatory Commission.
7. Institute of Electrical and Electronics Engineers (IEEE). (1997). *Guide for Incorporating Human Action Reliability Analysis for Nuclear Power Generating Stations, IEEE 1082.* New York: Institute of Electrical and Electronics Engineers.
8. Kolaczkowski, A., Forester, J., Lois, E., and Cooper, S. (2005). *Good Practices for Implementing Human Reliability Analysis (HRA), Final Report, NUREG-1792*. Washington, DC: U.S. Nuclear Regulatory Commission.
9. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
10. Swain, A.D., & Guttman, H.E. (1983). Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications. Final report. NUREG/CR-1278. Washington, DC: U.S. Nuclear Regulatory Commission.