Predicting Locomotive Crew Performance in Rail Operations with Human and Automation Assistance

Victoria Chibuogu Nneji, *Member, IEEE,* M. L. Cummings, *Senior Member, IEEE,* and Alexander J. Stimpson, *Member, IEEE*

***Abstract*—As new technologies are introduced into rail operations, models are needed to represent the task load of operators to identify periods of extreme workload that could be mitigated through technological interventions. To this end, a computational model is described to quantitatively simulate freight rail operator workload to understand the impacts of inserting intelligent automation on different crew configurations. A detailed task analysis served as the basis for identifying tasks performed during transit. Utilizing task characteristics and operating conditions as inputs, a discrete event simulation was designed to predict human operator workload. Results show that during heavy traffic conditions, the presence of automation can impact locomotive engineer performance more than the presence of a freight conductor in a short-haul freight rail setting. However, under typical conditions, assistance may not be as beneficial for human operator performance.**

***Index Terms*— human performance modeling, discrete event simulation, freight rail operations, automation, workload**

I. Introduction

THE U.S. rail transportation industry has a long history of introducing new technologies to meet evolving demands. Innovations such as advanced signaling technology and automated methods for tracking and distributing cars on complex networks of tracks have helped to improve safety by deterring many accidents. They have also facilitated smaller crew sizes and increased workload on crews [1].

Today, human factors are the leading causes of train accidents in the United States [2]. Despite this fact, rules proposed by the Federal Railroad Administration (FRA) in 2016 suggest that safety is significantly improved when the primary operator has another human in the locomotive [3]. Presumably, safety will also be positively impacted when the Congressional mandate set in 2008 for all major railroads to integrate automation in the form of positive train control (PTC) is fully enforced [4].

Simultaneously, companies like GE Transportation have been selling cruise control systems to become a standard for energy management, appealing to the US Environmental Protection Agency [5]. PTC essentially acts as a back-up monitor of unsafe track conditions and initiator of emergency braking, which is needed since humans are poor at monitoring tasks [6]. Combining such systems could provide a form of autopilot like those of self-driving cars. However, previous research in human-automation interaction has demonstrated that additional automation does not necessarily guarantee increased system effectiveness or safety [7], [8]. Often, automating a task within a larger system modifies the task by transferring the operator's workload from one physical or cognitive resource to another, thereby changing the task rather than eliminating one [9]. Poorly designed automation can contribute to errors and reduce system effectiveness due to implementations that increase cognitive workload.

Workload is of interest in rail operations as the job of human operators in transportation increasingly involves changing demands in cognition and decision making as manual labor is more often allocated to machines [10]. Cognitive workload is the "level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience" [11]. While the importance of workload is clear, numerous challenges remain in objectively measuring workload [12].

To this end, this research presents a task workload modeling approach commonly found in aviation settings, applied to freight rail operations. Developed using objective task time data collected during observations of train crews and analytical data collected from interviews with 18 subject matter experts (SMEs), the Simulator of Human Operator Workload (SHOW) represents locomotive crew workload under numerous operating conditions and task loads. SHOW extends previous efforts by developing a computational model to better understand how workload is affected by the introduction of automated technologies, as well as how these advanced technologies could or should inform crew configurations. The ultimate objective of SHOW is to provide various stakeholders with a tool to investigate potential staffing and technology architectures in conceptual design phases.

II. Literature Review

With increasing automation in the freight rail domain, it is important to understand how operator workload may be affected by introducing new technologies. Previous work in the rail domain includes the development of a hierarchical task network to abstractly represent workload [13]. Researchers have synthesized cognitive task analyses of engineers [14],

Manuscript submitted October 2017; revised April 2018 and October 2018; accepted December 2018. This work was supported by the U.S. Department of Transportation under Grant DTFR53115C00017. V.C. Nneji, M. L. Cummings, and A.J. Stimpson are with the Humans & Autonomy Lab of Duke Robotics, Duke University, Durham, NC 27708 USA (e-mails: vnneji2@gmail.com, m.cummings@duke.edu, ajstimps@gmail.com).

conductors [15], and dispatchers [16]–[18] and summarized the results in concept maps. They then verified the accuracy of the maps through interviews with subject matter experts (SMEs) and used the concept maps to develop an abstraction hierarchy [19], representing specific tasks and generalized functions executed by engineers and conductors for safe and effective rail transport.

This hierarchical task network provided valuable insight into the causes of operator workload. However, given the insertion of new technologies into rail operations, an additional modeling method is needed that allows for concept evaluation across different operating conditions, especially as such changes could significantly impact operator workload and system performance. Mathematical models represent one such analytical approach that can calculate human operator workload across different conditions and time scales [20].

Research in human supervisory control [21], human-machine interaction [22], air traffic control [23] have suggested time-on-task as a reliable estimation of workload. Time-on-task is defined as the duration an operator's attentional resources are actively used to meet functional requirements. Wierwille and Eggemeier [12] suggest that time-on-task estimations can reveal areas of heightened workload that may be missed in other forms of assessment. Cummings and Guerlain [24] and Donmez, Nehme, and Cummings [25] extended the use of task time data by incorporating a utilization metric as a means of objectively measuring high and low workload. This utilization metric is defined as percentage of time an operator spends on task performance out of total operation time. Maximum utilization is 100%, at which point there no additional capacity available for a human operator to allocate toward accomplishing tasks.

When utilization levels are too high, operators may be too busy to accumulate the information required to maintain situation awareness (SA). Similarly, when operators are underutilized, they could overlook information from the environment due to low engagement, also leading to poor SA. Levels of utilization below 30% have been associated with poor performance due to boredom and distraction while a 70% utilization threshold has been used to indicate the upper bound of optimal workload [26]–[29]. For over a century, researchers have used the Yerkes-Dodson inverted-U theory to conceptualize the relationship of workload levels to performance where performance declines during extreme workload periods of under- or over-utilization [25]. The goal for transportation systems in general, and rail operations specifically, is to optimally man crews such that each operator is moderately utilized.

Because of the temporal nature of shift work experienced by rail crews, we chose to develop a workload model through discrete event simulation (DES). DES has been used for years in modeling manufacturing [30], health care services [31], and military operations [32]. It allows designers to realize areas that cause undesired delays in their systems. However, in railway operations DES has only been used in some studies to model the utilization of infrastructure, without explicit consideration for human factors [33]–[35]. The rail industry needs an approach to modeling workload of train crew, with a special emphasis on the role of supporting humans with new technologies. With pending changes in operational practices, including the insertion of various forms of automation, stakeholders need a tool for rapid exploration of various crew workload and tasking models. Such a tool could answer questions like how is locomotive engineer performance affected by automation and how might performance change without a freight conductor?

This paper describes the design, validation, and analysis of a functional DES of freight rail operations to model crew workload called SHOW. Using DES is advantageous as it allows for a time-based representation of human processes at a task-level. Therefore, we began this modeling effort by collecting time-on-task data of real freight rail operations.

III. Empirical Data Collection

Cognitive task analyses were conducted through direct observation, and structured and unstructured interviews with SMEs aided in identification and categorization of in-cab crew tasks. Using present-day operations as a baseline, a typical in-cab crew can include a team of two or single-person operations, with the engineer as the primary operator. Based on a framework proposed by Subrahmaniyan et al. [13] that generalized functions to represent work performed by the engineer and conductor, a core set of tasks were defined as shown in Table I.

The work of the engineer is more than just maneuvering the locomotive control system. It also requires paying attention to the environment, in-cab displays, and radio communications with dispatch to maintain situation awareness and plan several miles in advance. On the other hand, the work of the conductor is to supervise the train conditions on the ground at terminal points and remain attentive to the engineer while the train is in motion in the case of an emergency. When the conductor engages in planning, monitoring inside, or exception handling tasks, they function as a backup to the engineer since tasks of those type are primarily allocated to the engineer who has been trained to handle them as a single operator.

We directly observed locomotive engineers for two 2.5-hour ride-alongs at a private Canadian regional railroad that operates with single-person crews. During the observations, our team noted a description and time for each task observed (e.g., blowing horn, 00:15:11-00:15:16). In total, approximately 50 tasks were logged per hour.

The ride-alongs were followed by structured interviews with our first set of seven SMEs: five in the railroad's operations, training, dispatch, and regulations along with two labor union representatives from a Canadian Rail United Transportation Union chapter. Experience across all personnel ranged from 3 to 20 years. Video recordings from ride-alongs were reviewed to identify times spent on each observable and partially-observable task performed throughout different phases of the shift (e.g., communicating, paperwork).

Interviews following observations provided clarification of tasks observed during normal railroad operations. At the Amtrak Training Center in Wilmington, Delaware, additional structured interviews were conducted with a second set of SMEs, four trained operators since their operations allow engineers to operate alone in the locomotive cab during short-haul shifts [36]. We observed engineer task performance in three physical locomotive cab simulators and one trainee class to supplement our dataset.

TABLE I. TYPES OF FREIGHT RAIL CREW TASKS

Task Type	*Description*
Monitoring Inside	Maintaining attentiveness to informational displays and the engineer's performance to maintain a safe operation
Monitoring Outside	Maintaining attentiveness to warnings and environmental conditions that may affect operations
Communicating & Coordinating	Filtering through relevant information for the operation and communicating information that may impact the macro-level network of operations
Paperwork	Reviewing and recording operating conditions
Exception Handling	Attending to unexpected or unusual situations that must be handled to continue with the trip
Planning Ahead & Generating Expectations	Maneuvering locomotive control system for throttle, braking and other subtasks in anticipation for oncoming dynamic motion requirements
Maintenance-of-Way Interactions	Maintaining situational awareness of other crews along the track
Temporary Speed Restrictions	Recalling information issued on track bulletins and adapting to updates while train is in motion
Signal Response Management	Maintaining attentiveness to direction from track signaling system and responsive to proper control system within a safely allotted time

One important result from these observations was the elucidation of different phases of operation. A phase is defined by a change in operational behavior which drives either different tasks for an operator, or substantially different frequencies of tasks. From the cognitive task analyses and observations, we identified three phases for freight rail operators: startup, full motion, and yard.

The startup phase includes the first 30 minutes of an engineer’s shift in the train. Startup includes tasks supporting regulatory requirements, such as communicating with dispatch and testing the emergency braking system so, just as in the real world, there is a chance that a task comes in last minute that takes longer than expected to complete.

During full motion the engineer drives the locomotive to its destination and this typically is the longest phase of the trip. The final phase, the yard phase, is the source of the highest rates of accidents [37].

IV. DESCRIPTIVE MODEL DEVELOPMENT

The results of these cognitive task analyses of the engineer and conductor functions were used to construct a descriptive temporally-based task model of crew responsibilities, including frequencies and durations. For this case study, we gathered estimates from an average of the observed times per task type under typical conditions and the “most likely” times reported by SMEs. For heavy traffic, we used the upper 3-quantile of times, including the “pessimistic” times reported by SMEs. We created a table of estimated time the locomotive engineer spent on each type of task for each 10-minute interval. These were aggregated to represent 5-hour short-haul shifts.

The results from this descriptive model were validated by subject matter experts for four defined scenarios that described workload of an engineer working alone or with a conductor during typical or high traffic conditions under both crew complements. One such scenario is presented in Fig. 1.

In Fig. 1, the scenario of an engineer operating alone under typical conditions is presented. Task times are described over 10-minute time ranges. The three markers labeled *Full Motion* beginning in the first 30 minutes, *Traffic* beginning at minute 200, and *Yard* approaching the final 30 minutes represent changes in tasks and frequencies of tasks. For example, the *Traffic* event represents the train approaching other vehicles on the track.

To better associate the nine task types in Table I with mission goals for freight rail that would ultimately be affected by operating conditions, we simplified the structure of engineer tasking to six. We organized Planning Ahead and Generating Expectations, Signal Response Management, Temporary

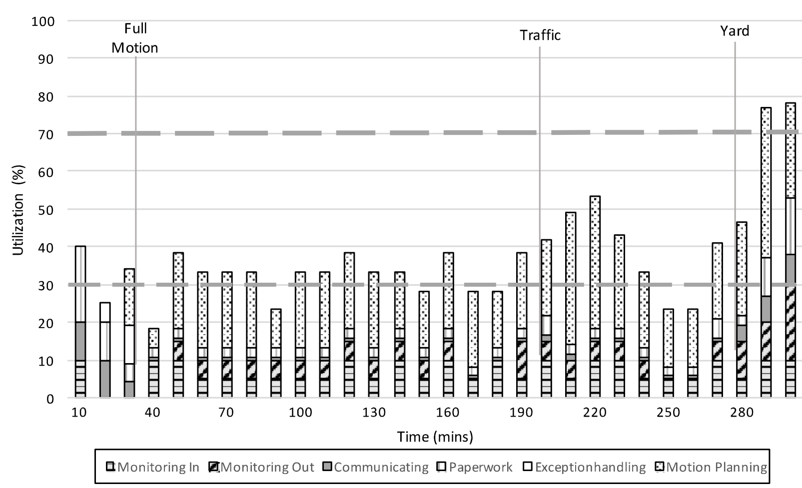

Fig. 1. Percent of engineer utilization in a one-person crew in 10-minute intervals over a 5-hour shift during the typical condition.

Speed Restriction Management, and Maintenance of Way Interactions under the higher-level task of Motion Planning. In Motion Planning, the engineer performs many lower level tasks to meet the higher-level goal of moving the train forward toward its destination efficiently. Motion Planning is unique to the engineer and cannot be directly reallocated to the conductor. The model shows that almost half, 46%, of the engineer's workload can be attributed to *motion planning*. Motion planning involves manipulating the locomotive control system and preparing in advance for mechanisms required to move the train towards a destination. Monitoring the displays within the cab makes *monitoring inside* an important secondary task that accounts for 22% of their total time on task during the 5-hour shift.

In this scenario that represents typical operations over a 5-hour period, typical of short-haul lines, the engineer is responsible for fulfilling all onboard tasks, and Fig. 1 shows that utilization ranges between 18% and 78%, with an average of 37% utilization. In this scenario, the engineer spends 20 minutes above 70% and 80 minutes below the 30% utilization threshold. With descriptive results, we generated data distributions and developed a predictive model in an iterative design and validation process depicted in Fig. 3 and further explained in the next section on SHOW.

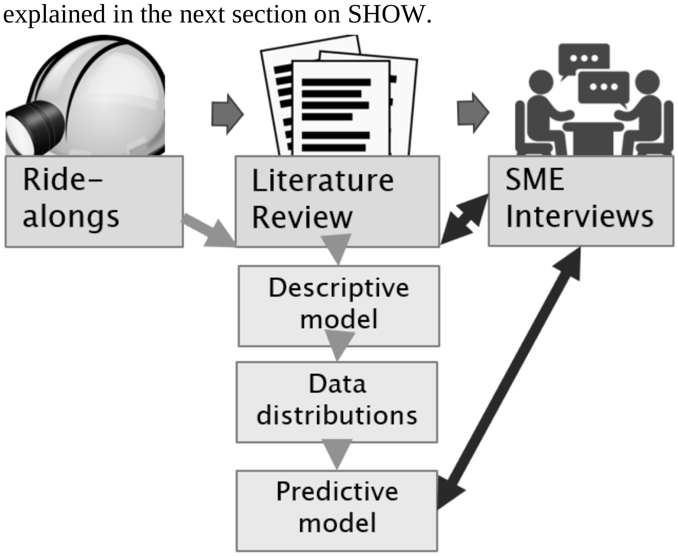

Fig. 3. Diagram of SHOW development and validation.

V. Simulator of Human Operator Workload

The Simulator of Human Operator Workload (SHOW) models human operators and assistive agents as serial processors of complex tasks and records the basic units of key performance metrics through the time it takes them to accomplish these tasks. The simulation begins when the first task arrives in the system (based on a stochastic arrival distribution) and then is assigned to an operator's queue based on the operator's functional capability to handle that type of task. Operators include humans (i.e. conductor and/or engineer) and automated agents (i.e. PTC and/or cruise control technologies). Tasks can queue, awaiting operator availability for servicing. In the process of service, the task may be interrupted by another task of higher priority [14] and thus returned to wait in queue due to preemptive priority scheduling, which models multitasking with rapid serial switching. Finally, the task exits when it is completely serviced or is expired.

The process of a task flowing through the freight rail simulation is illustrated in Fig. 2. At any point in this process, the task may expire before service, at which point it departs the system. A task may also drop if the phase changes or the trip ends while it is waiting in a queue. For example, at 30 minutes, no additional startup tasks can enter the queue and engineers must complete any tasks already in the queue before they can begin full motion phase.

The simulation allows tasks to have different arrival rates and expiration times by phase as appropriate. This is intended to mimic real-world operations, where types of tasks may differ between phases. For example, a maintenance of way interaction task would only arrive in the latter two phases of a trip. These task types were consistently categorized for both engineers and conductors. Tasks within each type of course vary and are represented by the random distribution of times.

The sources of system stochasticity are task inter-arrival and operator service times which are random observations generated from probability distribution functions. Each replication in a simulation run pulls from the random distributions, which simulates the variability of event times. Operator service times on tasks are simulated to increase by 1% each hour to model the impact of fatigue as demonstrated in the linear function from Hursh, Redmond, Johnson, Thorne et al. [38] that found reciprocal cognitive performance capacity

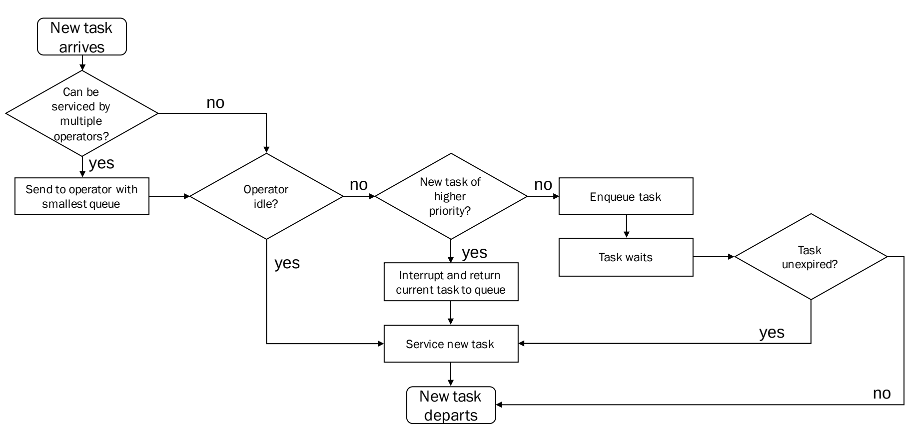

Fig. 2. SHOW process flow from task arrival to departure from freight rail operator system.

decays over consecutive hours of wakefulness.

More simulated trips provide a more robust set of results although the simulation's processing time will increase [39]. SHOW tracks several different statistics as shown in Table II, computed for each human operator across replications.

Utilization, the principal measure, is used as a proxy for workload. The simulation records the utilization for each operator throughout the shift and results are reported in 10-minute intervals. This output can be used to validate the simulation model output data when compared to the empirical data structure. The expired/completed tasks and average wait time outputs provide additional statistics to infer system performance metrics of effectiveness and efficiency.

SHOW is an online tool (*apps.hal.pratt.duke.edu/show*) that can be used by any interested party to model any rail platform, with the assumption that the modeler knows the tasks and at least basic estimates of time of task performance. The platform allows a person to input default tasks, times, and allocations to represent their own operational settings (Fig. 4). We generated default probability distributions from the whole set of original task arrival and service time estimates (Table III).

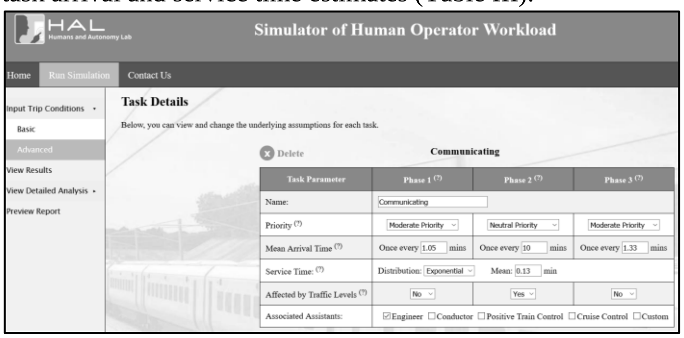

Fig. 4. Screenshot of SHOW user interface.

VI. Simulation Validation

The validation process establishes the model as a reasonable reflection of the real system it attempts to simulate [40]. First, the stand-alone simulation software SHOW, developed in C++, was verified with Rockwell Automation's off-the-shelf Arena version 14.7 package. Then a three-step validation process for our given short-haul freight rail case study, was followed:

1. Statistical comparison with descriptive model for external results validation
2. Subject matter expert review for face validation
3. Sensitivity analysis for internal validation

To validate the output from the overall simulation model, a Kolmogorov-Smirnov statistical goodness-of-fit test was performed to determine if its data closely compared to that of the actual system described. The test failed to reject the null hypothesis (α=.05) that the samples from the descriptive and predictive models were drawn from populations with identical distributions during the Full Motion phase of the shift. D=.2917 (< $D\alpha$), p = .216.

The second step involved gathering qualitative feedback from SMEs with experience in freight rail operations, which were seven freight rail original equipment manufacturers, researchers, and actual locomotive operators, both current and retired. The SME review qualitatively supported the model's results, including that conductors experienced overall lower workload than engineers. SMEs also attested to operators experiencing heightened workload during the final phases of shifts.

TABLE II. SIMULATION OUTPUT STATISTICS

Output Statistic	Description	Purpose
Utilization	Time on task divided by total (10 min) time interval	Workload measure
Expired/Completed Tasks	Tasks that were not versus tasks that were completed by the operator	Identify which tasks are more likely to be missed; throughput
Average Wait Time	Average time each task waited in the operator's queue	Identify which tasks spend the most time waiting

TABLE III. DEFAULT TASK ARRIVAL AND SERVICE TIME DISTRIBUTIONS

Task	***Phase***	***Arrival***		
		Rate (task /minute)	Traffic (0,1)	**Service Time (minutes)**
Monitoring Inside	Start Up	Exponential (**λ**=1/3.5)	0	Exponential (μ=.133)
	Full Motion	Exponential (**λ**=1/2.67)	0	
	Yard	Exponential (**λ**=1/2.67)	0	
Monitoring Outside	Start Up	Exponential (**λ**=1/9)	0	Exponential (μ=.15)
	Full Motion	Exponential (**λ**=1/5)	1	
	Yard	Exponential (**λ**=1/1.75)	0	
Communicating & Coordinating	Start Up	Exponential (**λ**= 1/1.05)	0	Triangular (min=.033, max=.2167, mode=.085)
	Full Motion	Exponential (**λ**=1/10)	1	
	Yard	Exponential (**λ**=1/1.33)	0	
Paperwork	Start Up	Exponential (**λ**=1/6.67)	0	Uniform (min=.05, max=1.5)
	Full Motion	Exponential (**λ**=1/20)	1	
	Yard	Exponential (**λ**=1/3.33)	1	
Exception-handling	Start Up	Deterministic (2x, by 10 and 20)	0	Lognormal (μ=.98, $\boldsymbol{\sigma}^2$=1.39)
	Full Motion	Exponential (**λ**=1/3000)	1	
	Yard	Exponential (**λ**=1/3000)	0	
Planning Ahead & Generating Expectations	Start Up	Exponential (**λ**=1/15)	0	Exponential (μ=.33)
	Full Motion	Exponential (**λ**=1/5)	1	
	Yard	Exponential (**λ**=1/2.5)	0	
Maintenance of Way (MOW) Interactions	Full Motion	Exponential (**λ**=1/600)	1	Uniform (min=.167, max=2.5)
	Yard	Exponential (**λ**=1/60)	0	
Temporary Speed Restriction	Full Motion	Exponential (**λ**=1/30)	1	Uniform (min=0, max=5)
Signal Response Management	Start Up	Exponential (**λ**=1/30)	0	Uniform (min=.5, max=2)
	Full Motion	Exponential (**λ**=1/10)	0	
	Yard	Exponential (**λ**=1/15)	0	

In general, those with careers directly involving freight rail operations found the results of both locomotive engineer and freight conductor to be like their own experiences. However, some researchers with experience studying workload in experimental controlled labs of physical cab simulators were surprised that utilization levels were reported to be lower than they imagined. So, the SMEs encouraged that we share the model with new stakeholders to expand our dataset and support SHOW's applicability to more freight rail operations since SHOW has the flexibility to computationally simulate over 2360 trillion settings of shift time, traffic level, and human- and automation-assistance that are too time-intensive and expensive to test in physical simulators with human volunteers.

Finally, a sensitivity analysis was performed to discover which model factors may significantly impact key performance indicators (KPIs). The goal of the analysis was to investigate how levels of variation in parameters impacted the deviation of average utilization, wait time, throughput rate, and expiration rate output measurements from the baseline results.

Our analysis revealed that variations in inter-arrival time significantly changed mean results for utilization and throughput rate. Variations in service time significantly changed mean results for utilization. And only extreme variations in traffic parameters significantly changed mean utilization and throughput rate. In contrast, no variations within +/-20% in the fatigue parameter significantly changed any KPIs. Overall, our three-step validation process built confidence in the model's real-world usefulness [41].

VII. Simulation Analysis

To demonstrate the utility of SHOW, we used the short-haul freight rail validated model in Fig. 1 to conduct a comparative analysis of operator workload for eight concepts of operations system reconfigurations. Our objective was to identify in the case of one engineer operating a locomotive, which combinations of human and automated assistants could yield better performance for the system.

Four levels of assistance and two types of conditions were defined (Table IV). With no assistance, the engineer is responsible for all tasks alone. Human assistance refers to an onboard freight conductor who can offload signal response management, monitoring inside, and planning tasks from the engineer. We defined automated assistance as integration of positive train control (PTC) and cruise control [42] technologies. Under the automated assistance concept, PTC could handle emergency braking, signal response, and monitoring inside tasks, while cruise control could handle planning ahead and generating expectations tasks with negligible service times. The *all* assistance category means that the engineer was supported both by the conductor and automation. For any shared tasks represented in SHOW, the engineer was the primary operator but if the engineer was occupied with a task and a new task arrived, the task was routed to the conductor's queue, and then to automation's queue if the conductor was unavailable.

Typical conditions represent *normal* operations as observed and validated. Heavy traffic conditions represent a *worst* day an operator may have with respect to traffic events, leading to overall high task load, defined as 3.5 times the typical task arrival rates and validated by the SMEs. A penalty function (1) was defined to identify how the locomotive engineer's performance may be influenced by levels of human and automation assistance during typical and heavy traffic conditions. Donmez et al. [25] found that extreme workload yielded inefficiencies in operator attention and degraded human supervisory control system performance. The thresholds for potential decrements in performance (penalties) are marked at 30% and 70% utilization based on results from previous experiments [26]–[29]. At those points, a penalty value (0,1] was computed along a slope of -1 for low workload and +1 for high workload.

Table IV. Design of Experiments

Experiment	*Levels*	
	Assistance	**Condition**
1	None	Typical
2	All	Typical
3	Automation	Typical
4	Conductor	Typical
5	None	Heavy Traffic
6	All	Heavy Traffic
7	Automation	Heavy Traffic
8	Conductor	Heavy Traffic

For each 10-minute interval recorded in the simulation, the following algorithm generates a penalty value according to the penalty function we define below:

$$\begin{aligned} &If\ Utilization < 30\%, \\ &\qquad Penalty = 1 - 3.33 * Utilization \\ &Else\ if\ Utilization > 70\%, \\ &\qquad Penalty = 3.33 * Utilization - 2.33 \\ &\quad Else, Penalty = 0. \end{aligned} \tag{1}$$

This penalty function (Fig. 5) approximates the Yerkes-Dodson theory which establishes performance as an inverted parabolic function of workload. In this inverse and linear approximation, penalty is a function of utilization (proxy for workload). The cost function assumed an equal linear penalty for low and high workload, as previous research [25] demonstrated that this approach effectively matched observed conditions.

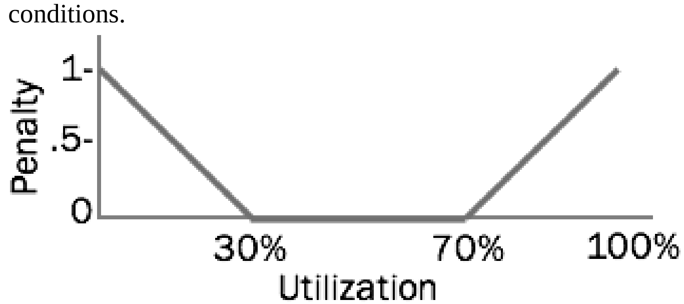

Fig. 5. The penalty function of operator utilization.

We applied this function across 1000 replications to analyze the distribution of the results for each of the eight experiments presented in Table IV. A one-way analysis of variance was used to determine whether the penalty values from extremely high workload under high traffic conditions differed when the locomotive engineer was alone, supported by all assistance, supported by automation only, or supported by human only (Fig. 6). Assumptions of normality and homogeneity of error variance were met.

The analysis showed significant differences among the four groups of 1000 data points, $F(3,3996) = 49.9$, $p < .001$. An engineer under high workload assisted with a combination of automation and human support exhibited the best performance in terms of penalty (M=2.69, SD=1.63), somewhat lower performance with automated assistance only (M=2.71, SD=1.61), poorer performance with human assistance only (M=3.21, SD=2.18), and worst performance without any assistance (M=3.55, SD=1.95). Post-hoc Tukey's honest significant difference tests showed that engineer performance during no-assistance operations differed significantly ($p < .001$) from each of the other three settings. The difference between all- versus automated-assistance was not statistically significant (p=.76).

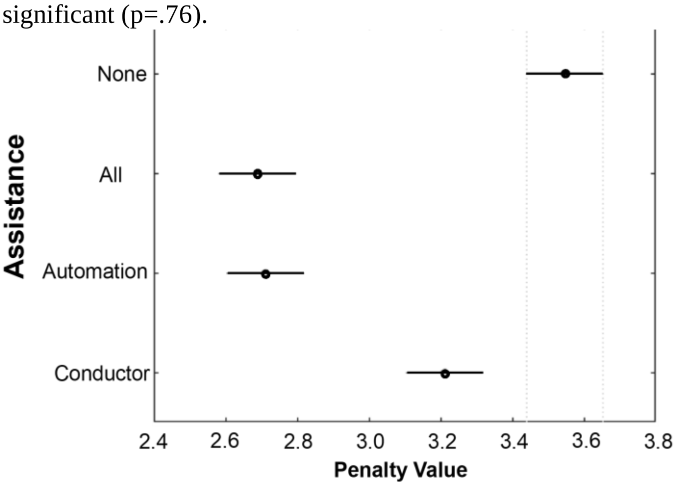

Fig. 6. Mean penalty values with comparison intervals (α=.001) for high workload experienced by engineer in heavy traffic condition with each level of assistance. "None" level of assistance was significantly different from every other level.

While, on average, all assistance led to better performance under the high workload condition, the simulation results show that the presence of automation with or without the conductor made much more of a difference ($p < .0001$) than with the conductor alone (p=.0003).

However, under typical low workload conditions, any form of assistance led to significantly higher penalty values as shown in Fig. 7. A one-way analysis of variance was used to determine whether penalty values differed when the locomotive engineer was alone, supported by all assistance, supported by automation only, or supported by human only. Assumptions of normality and homogeneity of error variance were met.

The analysis showed significant differences among the groups of data. $F(3,3996) = 75.77$, $p < .001$. A lone engineer exhibited the best performance based on the penalty score (M=10.25, SD=1.48), somewhat lower performance with human assistance (M=10.96, SD=1.48), poorer performance with automation (M=11.06, SD=1.41), and worst performance when assisted with both human and automation (M=11.09, SD=1.41). Post-hoc Tukey's honest significant difference tests showed that engineer performance during no-assistance operations differed significantly ($p < .001$) from each of the other three configurations, but the differences between all- versus automation (p=.69), all- versus human-assistance (p=.04), and automation versus human-assistance (p=.10) were not statistically significant. So, in simulated operations, under typical low task loading, engineers working alone performed better than with any assistance, while under heavy traffic conditions, engineers working with automation (with or without human assistance) had the best performance of the tested conditions.

Interestingly, the average penalty acquired under typical low workload conditions (M=10.84) was significantly higher than under high workload conditions (M=3.04). In referring to the penalty function (1), this may be attributed to extended periods of time at or below 30% utilization during the shift. It should be emphasized that these interpretations are strictly just for this set of operations and the model would need to be calibrated for each specific application.

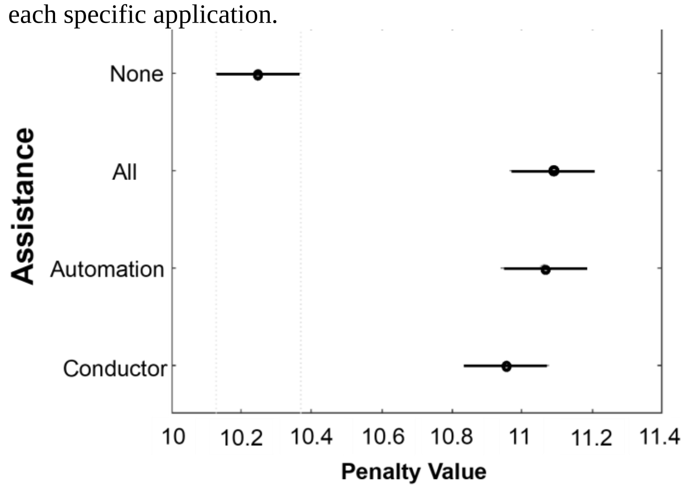

Fig. 7. Mean penalty values with comparison intervals (α=.001) for low workload experienced by engineer in typical conditions with each level of assistance. "None" level of assistance was significantly different from every other level.

VIII. Discussion & Conclusion

SHOW was developed as a discrete event simulation to model rail operators in various operating conditions to investigate various staffing and technology architectures. We gathered empirical data from the railroad industry and validated a short-haul freight rail model using SHOW through statistical goodness-of-fit tests, subject matter expert review, and sensitivity analyses.

As with any model, there are several limitations based on the scope of our analyses. First, SHOW only models the specific tasks and traffic densities selected by the modeler, including the associated distributions and parameters. Thus, any result will not accurately reflect all freight rail scenarios. Second, the model does not yet incorporate the possibility for human error in processing tasks. This addition is currently in development. Third, we did not model possible new tasks generated from coordinating with human and/or automation assistance. The simulation needs predictive validation before decisions can be more confidently drawn from the prospective analysis presented here. Finally, the model does not account for some human factors such as the hours of sleep prior to the shift that could impact performance. Investigating how robust the homeostatic linear representation of cognitive decay due to fatigue across different rail operating conditions would be a useful contribution.

SHOW is not meant to be a fine-grained model of an individual's response, rather a systems-level response for planning future system architectures. Comparing performance across alternative short-haul system configurations in SHOW, we found that the impact of automation on performance may be more helpful than a freight conductor at moderating locomotive engineer workload during heavy traffic conditions. However, that same automation could be detrimental to operator performance in a typical short-haul freight rail scenario.

With companies already installing automation and making decisions about the future workforce [43], a simulation model like SHOW can aid in the investigation of potential implications of human- and automated- assistance on the primary operator's workload across operating conditions. Particularly in the setting of freight rail operations, as automation increases, we expect boredom and the associated reductions in human performance to also increase [27], [44]. PTC, combined with other technologies like energy management cruise control systems can provide efficiencies that lead many railroads to believe that the second crewmember is unnecessary [45]. The results from this application of SHOW to a short-haul case study demonstrate that we cannot take a blanket approach to improving safety conditions. Further consideration is required to identify when and how adaptive levels of automation could be implemented to modulate workload and optimize human performance across different phases and operational conditions.

Future research is needed to better understand the unexpected challenges that different types of automation and human assistance may bring to a primary operator's performance. For example, the presence of assistance may lead to new tasks that we failed to model in this first iteration. Moreover, given increasingly capable automated systems, another consideration is shifting the engineer off-board to remotely control one or more trains at a supervisory level. When thoughtfully designed, human supervisory control has been shown to be beneficial to system performance in similar operational settings with adaptive automation [46].

Limitations were identified in our approach, but we designed SHOW to allow any person to perform their own simulation analysis with any railroad operations even if operators are modeled to handle a specific set of tasks. The validation strategy demonstrated that SHOW is a potentially useful tool for modeling operator workload in freight rail environments that can simulate operations to allow for study of potential future system changes.

ACKNOWLEDGMENT

Our views do not necessarily reflect views of the US DOT. Dr. Michael Clamann and Branch Vincent helped in the data collection and simulation development phases. We thank the subject matter experts who shared time and wisdom with us.

REFERENCES

[1] C. D. Martland, "Workload Measurement and Train Crew Consist Adjustments: The Boston & Maine Experience," *Transp. J.*, vol. 21, no. 3, pp. 34–57, 1982.

[2] FRA, "Summary of Accident/Incident Counts from January-December 2015 Priliminary Data," *U.S. Department of Transportation Research*, 2016. [Online]. Available: http://safetydata.fra.dot.gov/officeofsafety/publicsite/summary.aspx. [Accessed: 01-Jan-2016].

[3] FRA, *Notice of Proposed Rule Making: Train Crew Staffing*. 2016, pp. 1–205.

[4] 110th US Congress, *Rail Safety Improvement Act (RSIA)*. 2008, p. Public Law 110-432.

[5] A. Brecher and M. Shurland, "Study on Improving Rail Energy Efficiency (E2): Best Practices and Strategies," in *Joint Rail Conference*, 2015, pp. 1–8.

[6] D. B. Kaber and M. R. Endsley, "Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety," *Process Saf. Prog.*, vol. 16, no. 3, pp. 126–131, 1997.

[7] R. Parasuraman and V. Riley, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 39, no. 2, pp. 230–253, 1997.

[8] M. L. Cummings and J. Ryan, "Shared Authority Concerns in Automated Driving Applications," *J. Ergon.*, vol. 3, pp. 1–4, 2013.

[9] A. Sebok, C. D. Wickens, and L. Laux, "Human-Automation Interaction : Lessons Learned from the Aviation Industry and Recommendations to the Rail Industry for Regulatory Guidance," Washington, D.C., 2015.

[10] W. T. Singleton, *The mind at work: Psychological ergonomics.* 1989.

[11] M. S. Young and N. A. Stanton, "Mental Workload," in *Handbook of Human Factors and Ergonomics Methods*, CRC Press LLC, 2005, pp. 391–399.

[12] W. W. Wierwille and F. T. Eggemeier, "Recommendations for Mental Workload Measurement in a Test and Evaluation Environment," *Hum. Factors*, vol. 35, no. 2, pp. 263–281, 1993.

[13] N. Subrahmaniyan, A. Liu, B. Miller, H. Groshong, and J. D. Brooks, "Investigation of New Roles for Humans and Automation in Rail : An Interim Report," Washington, D.C., 2014.

[14] E. Roth and J. Multer, "Technology Implications of a Cognitive Task Analysis for Locomotive Engineers," Washington, D.C., 2007.

[15] H. Rosenhand, E. Roth, and J. Multer, "Cognitive and Collaborative Demands of Freight Conductor Activities Results and Implications of a Cognitive Task Analysis," Washington, D.C., 2012.

[16] E. M. Roth, N. Malsch, and J. Multer, "Understanding How Train Dispatchers Manage and Control Trains: Results of a Cognitive Task Analysis," Washington, D.C., 2001.

[17] L. Huang, M. Cummings, and V. Nneji, "Preliminary Analysis and Simulation of Railroad Dispatcher Workload," in *Human Factors and Ergonomics Society 62nd Annual Meeting*, 2018.

[18] V. C. Nneji and M. L. Cummings, "Toward Level 5: What Autonomous Vehicle Companies Can Learn from Railroad and Airline Operations," in *Transportation Research Board 98th Annual Meeting (accepted)*.

[19] K. J. Vicente, "Cognitive Work Analysis," *Analysis*, vol. 17, no. 3, pp. 313–21, 1999.

[20] C. Rusnock, B. Borghetti, and I. McQuaid, "Objective-Analytical Measures of Workload – the Third Pillar of Workload Triangulation?," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9183, pp. 124–135, 2015.

[21] M. L. Cummings and C. E. Nehme, "Modeling the impact of workload in network centric supervisory control settings," *Neurocognitive Physiol. factors Dur. high-tempo Oper.*, pp. 23–40, 2010.

[22] W. B. Rouse, "Models of human problem solving: Detection, diagnosis, and compensation for system failures," *Automatica*, vol. 19, no. 6, pp. 613–625, 1983.

[23] D. K. Schmidt, "A queuing analysis of the air traffic controller's work load," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6. pp. 492–498, 1978.

[24] M. L. Cummings and S. Guerlain, "Developing operator capacity

estimates for supervisory control of autonomous vehicles.," *Hum. Factors*, vol. 49, no. 1, pp. 1–15, 2007.
[25] B. Donmez, C. Nehme, and M. L. Cummings, "Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control," *IEEE Trans. Syst. Man. Cybern.*, vol. 40, no. 6, pp. 1180–1190, 2010.
[26] R. M. Yerkes and J. D. Dodson, "The Relation of Strength of Stimulus to Rapidity of Habit-Formation," *J. Comp. Neurol. Psychol.*, vol. 18, no. 5, pp. 459–482, 1908.
[27] M. L. Cummings, F. Gao, and K. M. Thornburg, "Boredom in the workplace: A new look at an old problem.," *Hum. Factors*, vol. 58, no. 2, pp. 279–300, 2016.
[28] M. L. Cummings, C. Mastracchio, K. M. Thornburg, and A. Mkrtchyan, "Boredom and distraction in multiple unmanned vehicle supervisory control," *Interact. Comput.*, vol. 25, no. 1, pp. 34–47, 2013.
[29] W. B. Rouse, *Systems engineering models of human-machine interaction*, vol. 6. North-Holland, 1980.
[30] Y. W. Ng, J. Chan, and S. Kamaruddin, "Analysis of Shop Floor Performance through Discrete Event Simulation : A Case Study," *J. Ind. Eng.*, vol. 2014, pp. 1–10, 2014.
[31] V. Ahalt, N. T. Argon, S. Ziya, J. Strickler, and A. Mehrotra, "Comparison of emergency department crowding scores: a discrete-event simulation approach," *Health Care Manag. Sci.*, pp. 1–12, 2016.
[32] D. K. Mitchell, "Mental Workload and ARL Workload Modeling Tools," 2000.
[33] M. H. Cha and D. Mun, "Discrete event simulation of Maglev transport considering traffic waves," *J. Comput. Des. Eng.*, vol. 1, no. 4, pp. 233–242, 2014.
[34] C. Woroniuk and M. Marinov, "Simulation modelling to analyse the current level of utilisation of sections along a rail route," *J. Transp. Lit.*, vol. 7, no. 2, pp. 235–252, 2013.
[35] A. Nash and D. Huerlimann, "Railroad simulation using OpenTrack," *Comput. Railw. IX*, pp. 45–54, 2004.
[36] Oliver Wyman Group, "Analysis of North American Freight Rail Single-Person Crews : Safety and Economics Association of American Railroads," Boston, MA, 2015.
[37] FRA, "Federal Railroad Administration Status Update on Positive Train Control Implementation," Washington, D.C., 2016.
[38] S. R. Hursh *et al.*, "Fatigue models for applied research in warfighting," *Aviat. Space. Environ. Med.*, vol. 75, no. 3, Suppl., pp. A44-53, 2004.
[39] S. Robinson, *Simulation: the practice of model development and use*. 2004.
[40] A. M. Law and W. D. Kelton, "Simulation Modeling and Analysis," *ACM Transactions on Modeling and Computer Simulation*, vol. 2, no. 2. pp. 1370–1375, 2000.
[41] F. Leal, R. F. D. S. Costa, J. A. B. Montevechi, D. A. De Almeida, and F. A. S. Marins, "A practical guide for operational validation of discrete simulation models," *Pesqui. Operacional*, vol. 31, no. 1, pp. 57–77, 2011.
[42] C. R. Cole, S. Simson, M. McClanachan, T. McLeod, and J. Phelan, "Train cruise control," in *In proceeding of the 9th International Heavy Haul Conference*, 2009, pp. 638–645.
[43] G. Peters, "Are driverless freight trains safe?," *Railway Technology*, 20-Sep-2016.
[44] D. Sussman and M. Coplen, "Fatigue and Alertness in the United States Railroad Industry Part I: The Nature of the Problem," *Transp. Res. part F traffic Psychol. Behav.*, vol. F, no. 48, pp. 1–8, 2001.
[45] B. Stephens, "CSX's Ward says one-person crews are 'inevitable,'" *Trains Magazine*, 2017. [Online]. Available: http://trn.trains.com/news/news-wire/2017/01/18-csx-ward-one-person-crew. [Accessed: 21-Oct-2017].
[46] C. A. Miller and R. Parasuraman, "Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control," *Hum. Factors*, vol. 49, no. 1, pp. 57–75, 2007.

Victoria Chibuogu Nneji (M'16) is a creative strategist and strategic creator passionate about mobility. She graduated with a Master of Engineering Management at Duke University and B.S. in Applied Mathematics at Columbia University in NYC. As a PhD candidate in Duke Robotics, her work spans trains, planes, and automobiles.

Mary (Missy) Cummings (SM'03) received her Ph.D. in Systems Engineering from the University of Virginia in 2004. She is a professor in the Duke University Institute of Brain Sciences, Electrical and Computer Engineering, and Computer Science Departments. She is the director of Duke Robotics and the Humans and Autonomy Laboratory.

Alexander Stimpson (M'14) received his B.S. in biological engineering from the University of Florida in 2007, and the S.M. in Aeronautics and Astronautics from the Massachusetts Institute of Technology in 2011. His dissertation focused on the application of machine learning models to inform training assessment and intervention.