The Next Generation of Fatigue Prediction Models: Evaluating Current Trends in Biomathematical Modelling for Safety Optimization

Micah K. Wilson[1†], Luke Strickland[1†], Timothy Ballard[2], & Mark A. Griffin[1]

[1]Future of Work Institute, Curtin University, Perth, Australia

[2]University of Queensland, Brisbane, Australia

Full Publication Note

Please cite this article as: Wilson, M. K., Strickland, L., Ballard, T., & Griffin, M. A. (2022). The next generation of fatigue prediction models: Evaluating current trends in biomathematical modelling. *Theoretical Issues in Ergonomics Science*, *0*(0), 1–23. https://doi.org/10.1080/1463922X.2022.2144962

Author Note

†Luke Strickland and Micah K. Wilson contributed to the work equally. This research was supported by funding from the Australian Defence Science and Technology Group (MyIP:9079) and by a Forrest Research Foundation Prospect Fellowship awarded to Micah Kate Wilson. Address correspondence to Micah Kate Wilson, Curtin University, 78 Murray Street, Perth, Western Australia 6000 (micah.d.wilson@curtin.edu.au).

Open Science Note

Full information pertaining to datasets used in the manuscript, including the code used to conduct modelling and create plots, are provided in the repository associated with this pre-print accessible via: https://osf.io/yurvx/

Abstract

Biomathematical models (BMMs) are parametric models that quantitatively predict fatigue and are routinely implemented in fatigue risk management systems in increasingly diverse workplaces. There have been consistent calls for an improved "next generation" of BMMs that provide more accurate and targeted predictions of human fatigue. This review examines the core characteristics of next-generation advancements in BMMs, including tailoring with field data, individual-level parameter tuning and real-time fatigue prediction, extensions to account for additional factors that influence fatigue, and emerging nonparametric methodologies that may augment or provide alternatives to BMMs. Examination of past literature and quantitative examples suggests there are notable challenges to advancing BMMs beyond their current applications. Adoption of multi-model frameworks, including quantitative joint modelling and machine learning, was identified as crucial to next-generation models. We close with general recommendations for researchers and model developers, including focusing research efforts on understanding the cognitive dynamics underpinning fatigue-related vigilance decrements, applying emerging dynamic modelling methods to fatigue data from field settings, and improving the adoption of open scientific practices in fatigue research.

Keywords: fatigue, biomathematical model, alertness, neurobehavioral performance, cognitive modelling

Declarations of interest: None

1 Introduction

Fatigue is often defined as a physiological state of reduced mental or physical performance capability resulting from sleep deprivation, circadian processes, or other situational factors (Noy et al., 2011). In situations where failures of sustained vigilance can have serious consequences, fatigue prediction is often implemented to mitigate risk. Biomathematical models (BMMs) are often applied to predict the neurobehavioral outcomes of fatigue (e.g., alertness or response time) using time of day and sleep/wake history (for an overview, consult Civil Aviation Safety Authority, 2014). For example, airlines utilize crew management systems that coordinate workforce allocation across the globe using projected fatigue, and militaries utilize BMMs to implement watchkeeping schedules that optimize operational readiness. The proliferation of BMM tools (e.g., Hursh et al., 2004; Roach et al., 2004) has supported fatigue management in safety-critical work domains such as aviation, transportation, construction, and defence. In these contexts, practitioners typically predict fatigue using pre-configured 'default' BMM implementations that provide population average fatigue forecasts. Such implementations have several applications, for example they help to compare the relative fatigue risks of alternative work rosters, facilitate the design and planning of future technical systems (Boeing et al., 2020), and support accident investigation procedures (Price & Coury, 2015).

Due to the success of BMMs, the increasing abundance of data in modern workplaces, and the rise of increasingly powerful automation technologies, there have been calls to develop new fatigue prediction methods with additional capabilities (Dawson et al., 2011; Gunzelmann et al., 2019). We refer to these desired advancements as *next-generation fatigue modelling*, consistent with prior literature (e.g., Dawson, 2012; Dawson et al., 2011; Stone et al., 2020). Although next-generation fatigue prediction could be developed using a range of approaches, we focus primarily on development of models that are extensions or adaptations to BMMs. This is because BMMs are routinely employed in fatigue-risk management systems in industry applications and continue to receive significant research interest. Where relevant, we discuss complementary non-parametric approaches to fatigue prediction (2.4) and alternative approaches for latent variable estimation (3.1). The need for improved BMMs has been recognized for some time by both industry and researchers (Flight Safety Foundation, 2005; Hursh et al., 2004; Klerman & Hilaire, 2007; Reifman, 2004) and the core limitations of existing BMMs have been previously reviewed (Dawson, 2012; Dawson et al., 2011). The most thoroughly researched approaches to improve BMMs include tailoring them to match the fatigue dynamics of work environments and populations of interest; individualizing them to specific operators or individuals (Liu et al., 2017; Reifman et al., 2007; Van Dongen et al., 2012); and expanding them to incorporate a wider range of fatigue-related factors, such as workload (e.g., Honn et al., 2016; H. T. Peng et al., 2018). The central theme across these advancements is a need for more accurate and targeted predictions of human fatigue. Such models would have significant implications for safety-critical job domains in which teams must contend with significant environmental and workplace demands (e.g., time pressure, confinement, danger), and maintain high levels of vigilance, performance, and safety over lengthy missions (Bell et al., 2016; Bishop, 2004; Mallis & DeRoshia, 2005).

Enthusiasm for next-generation BMMs remains high (e.g., Bendak & Rashid, 2020; Civil Aviation Safety Authority, 2014; Flynn-Evans et al., 2020; Stone et al., 2020) and is stimulated by growth in fatigue science, advances in artificial intelligence, and emerging sensor technologies that detect physiological fatigue responses. Despite this enthusiasm, and substantial research efforts, progress is still in early stages. Research is limited primarily to experimental proof-of-concepts, with few next-generation features validated in or applied to the industries where predictive improvements are most crucially needed. Further, calls for next-generation models have echoed throughout scientific and industry-focused publications since the early 2000s, yet there remains a scarcity of successful implementations. In 2004, Dinges concluded that “Most current models of fatigue and its effects on performance appear to be more descriptive curvefitting, than theoretically driven, hypothesis-generating, data-organizing, mathematical approaches” (p. A182). There have been few changes in this regard in the years since.

A pressing question of concern is why has this research plateaued? Are there barriers, such as statistical constraints, that have slowed down the enhancement of BMMs and their application to relevant industries? In this article, we aim to describe the limitations of current methods and stimulate new avenues of research and development. In doing so, this paper also serves to consolidate the heterogeneous research on fatigue prediction into a more complete analysis of current development and progress, including emerging methods that can support a better understanding of the dynamics of fatigue. Elucidating the limits of BMMs does not preclude their continued use or refinement, instead it improves the certainty practitioners and researchers can have regarding their realistic effectiveness, in turn, fostering new avenues of research and safety optimization.

We begin the paper by reviewing the key characteristics of next-generation models, focusing on tuning model parameters using field-derived data (2.1), individual-level parameter tuning and real-time fatigue prediction (2.2), and extensions to modelling algorithms to account for additional factors that influence fatigue (2.3). We then review alternative emerging methodologies in statistical learning and artificial intelligence which may augment or provide alternatives to BMMs (2.4). Throughout these sections, we summarize recent research progress, identify practical and theoretical constraints limiting real-world applications, and where appropriate utilize simulations and use quantitative examples to explicate our arguments. We conclude the paper with a general summary of our findings, and discuss key challenges and opportunities facing the field of fatigue science.

2 Next-generation Fatigue Prediction Methods

Next-generation models offer opportunities to extend the applicability of fatigue prediction beyond their current capabilities of forward scheduling and population-average roster analysis. Three capabilities are core to the propositions of how BMMs can be adapted to meet next-generation goals. The first is that next-generation BMMs should be *tailored* appropriately to work populations and contexts of interest. At present, current generation BMMs are largely developed in tightly controlled laboratory settings, with research samples of convenience. Thus, there is general consensus that BMMs lack extensive validation in many operational contexts and bear only a coarse relationship with real-world risk (Dawson, 2012; Dawson et al., 2017; Gander et al., 2011; James et al., 2018; Reifman et al., 2007; Riedy et al., 2020). The second capability is an important related case of model tailoring known as individualisation — that is, the capability to *individualize* predictions to specific employees of interest. Many fatigue prediction scenarios, such as identifying the risk of nonoptimal performance or human error in a work environment, require specific predictions about the performance of each operator. Unsurprisingly, current generation BMMs, that focus on group-level average predictions, have proved to be poor predictors of individual-level performance in the field (e.g., in simulated lunar habitation see, Flynn-Evans et al., 2020; in naval submarine activities see, Wilson et al., 2021). The third capability is that next-generation BMMs should be *extended* to incorporate additional fatigue-relevant factors into projections. Current research has focused predominantly on the influence of pharmaceutical fatigue counter-measures (Ramakrishnan et al., 2013), chronic sleep dept (Rajdev et al., 2013), and task demands or workload (Honn et al., 2016; H. T. Peng et al., 2018). Though not a core capability, the incorporation of non-parametric methods has been argued to be essential to the realisation of next-generation models (Reifman, 2004). Below, we detail each of these key areas further. In doing so, we directly assess the limitations of BMMs for next-generation fatigue modelling and also provide readers with practical directions for future research.

2.1 Tailoring BMMs to Work Populations and Contexts

A frequently raised concern about current generation BMMs is that the laboratory conditions under which they are developed do not accurately represent the fatigue dynamics that occur in the work populations and scenarios of application (Dawson et al., 2017; Dean et al., 2007; Williamson et al., 2011). Indeed, many operational contexts involve challenges that make the assumptions of default BMM parameterizations inappropriate. For instance, in the submarine context, the lack of exposure to natural light sources in conjunction with artificial sleep-wake patterns (demanded by rostering constraints) is thought to disrupt circadian processes and rhythmicity, potentially altering the predictive contribution of the circadian and ultradian processes in BMMs (Cham et al., 2021; Guo et al., 2020; Sandal et al., 2006). Similarly, sleep quality can be disrupted by environmental factors such as motion or ambient noise, potentially influencing the homeostatic recovery rate (Beare et al., 1981; Guo et al., 2020). Thus, the use of BMMs estimated based on laboratory data may limit the accuracy of workplace fatigue predictions. In turn, this limitation may compromise risk mitigation efforts when performing forward-scheduling or mission planning (Flynn-Evans et al., 2020; Reifman et al., 2007; Wilson et al., 2021).

There are several ways that BMMs could be augmented or adjusted to represent the dynamics of a work population of interest more appropriately. One conceptually straightforward method is parameter tuning. BMMs include free parameters, which theoretically can index variations in fatigue dynamics across individuals or work contexts (Van Dongen et al., 2007). It follows, that a straightforward way to ‘tailor’ or ‘tune’ a BMM is to adjust model parameters to better describe the observed fatigue measurements from a work setting of interest. In later sections, we will discuss more complex approaches, such as augmenting the actual underlying BMM formulae to take account of additional variables. However, the importance of tuning the parameters of current generation BMMs can reveal important statistical constraints underlying the models.

Tailoring BMMs to specific populations or contexts requires representative fatigue data that models can be trained and parameterized with. Obtaining appropriate fatigue data that improves BMM predictions, relative to using default parameterizations, is challenging. One solution is to conduct high-fidelity laboratory studies with a relevant workplace population sample and simulate their expected workplace conditions in terms of tasks and work composition, for example with the use of synthetic task environments (Flynn-Evans et al., 2020; Gonzalez et al., 2005). This approach allows for exposure to natural light to be experimentally manipulated, the timing and duration of sleep periods, or timing of work-representative tasks (and workload), to be controlled. This method dramatically improves external validity while retaining the experimental control required for model estimation and development (e.g., Vital-Lopez et al., 2021).

Applying an experimental approach may not be feasible in many industries due to the constraints associated with how accurately the relevant factors affecting fatigue in a workplace can faithfully represented with laboratory resources. For many industries, it can be costly to retain experts in laboratory studies for the durations necessary to ascertain fatigue trajectories, which may be on the order of several days (e.g., in maritime domains, van Leeuwen et al., 2020). Therefore, an appealing alternative is to capture relevant fatigue and sleep data directly in the work environment (i.e., measure individuals in the operational context). The prospective benefits of this approach are: 1) it inherently offers the best chance of capturing the strain and recovery dynamics directly as they unfold in response to the environmental stressors that influence the underpinning neurobiological fatigue process; and

2) field fatigue measurement is essential for ‘real time’ prediction, in which future fatigue forecasts are updated based on incoming observations of field data. In this way, efforts to improve field measurement can concurrently support the development of methods which do not require all fatigue data to be collected prior to prediction.

There are challenges to this approach that must be considered. To successfully estimate BMM parameters from field data representative of important work contexts, data must be minimally invasive to collect, and yet comprehensive enough to identify the complex non-linear dynamics specified by BMMs. Given there are many possible processes that may underlie fatigue in the field, it is probable that fatigue and sleep measurement is subject to significant noise. If field measurements are too sparse or of insufficient quality to provide reliable estimates of true underlying fatigue dynamics, a BMM trained on that sample could fail when used to predict new data. Cross validating trained BMMs on new data can provide some level of assurance (e.g., Ramakrishnan et al., 2016). However, cross validation does not speak to the ability of BMMs to *measure* the fatigue processes of an individual or group, a goal which has been pursued in the literature (Ramakrishnan et al., 2015). It speaks only to predictive accuracy for a particular set of data, and therefore provides no guarantees that a set of BMM parameter estimates will generalize to data ranges (e.g., sleep schedules) outside of those that have been tested. Thus, a key step in determining the feasibility of tailoring BMMs is to understand their estimation properties in simulations with field-like data.

In the following section, we explore the feasibility of estimating a BMM using field data with a simulated *parameter recovery* study. With this approach, estimation properties are interrogated by simulating data from a set of known parameter values, then (treating the synthetic data as if it were real data) applying an estimation technique and checking the extent to which estimated values match the true values. Parameter recovery has been called for in the fatigue science literature (e.g., Reifman et al., 2007) and can inform the capability of BMMs to meet next-generation needs. To foreshadow, our results indicate that under highly favourable assumptions (regarding sampling frequency, measurement accuracy, and underlying fatigue dynamics) some model parameters can be well-estimated from field data, but important parameters relating to the homeostatic process are poorly estimated.

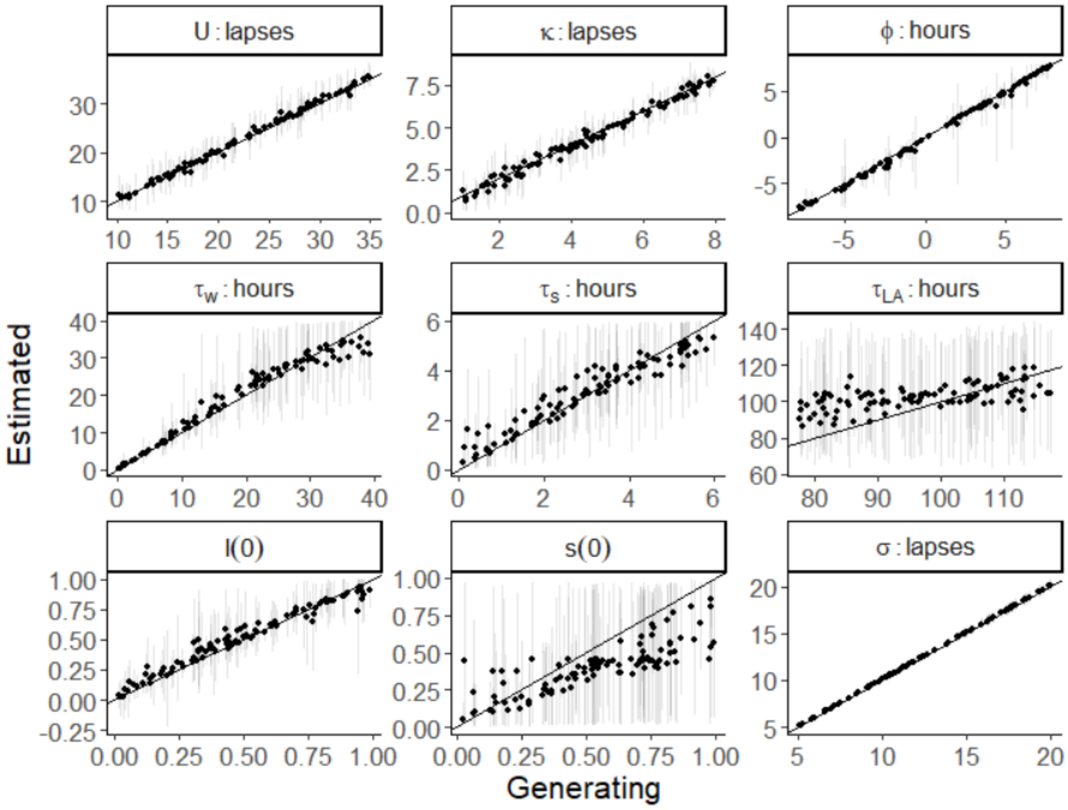
2.1.1 Parameter Recovery Study

The parameter recovery data structure is derived from an intensive longitudinal study of 64 navy submariners, across three submarine activities which lasted from 8 to 12 days each (see Wilson et al., 2021, for further details). Compliance was high, with the protocol embedded in work routines. Thus, we believe the data are of the upper bound of quality for a field scenario without risking extraneous demands to submariners. We generated a simulated dataset that matched the actual data with respect to sleep/wake patterns and fatigue observation timing and frequency (N = 1749). Further details of the measurement protocol, data structure, and model fitting procedure are included in the supplementary materials.

We examined the parameter recovery properties of the “unified model of performance” (Rajdev et al., 2013) because it is analytically tractable and it includes a sleep debt mechanism that theoretically accounts for the chronic fatigue accumulation likely to occur in operational environments (Liu et al., 2017; Rajdev et al., 2013; Ramakrishnan et al., 2013). 100 different sets of unified model parameter values were sampled (see supplementary materials for parameter ranges and sampling approach). Note that some BMM parameters depend on the scale of the outputted prediction (e.g., psychomotor vigilance task [PVT] mean response time, PVT lapses). We scaled the outputted fatigue prediction to approximately cover the number of lapses expected on a 10-minute PVT in order to match the BMM literature (e.g., Rajdev et al., 2013). For each of the 100 parameter sets, we simulated fatigue data from the model (with the respective parameter set), using the true recorded submariner sleep/wake patterns and fatigue measurement timestamps as inputs. We then fit the unified model to this simulated data in order to obtain the *recovered* parameter estimates. A match between the recovered parameter values and the original generating parameters (ground truth) indicates identifiability (i.e., good parameter recovery) — which is, the extent that parameters unambiguously describe the observed data better than any other set of parameters.

We used Stan (Carpenter et al., 2017) for the R Language (R Core Team, 2020) which estimates parameters with Bayesian Markov Chain Monte Carlo methods. This provides information not only about the most likely parameter estimates, but also the distribution of possible values, thus capturing uncertainty. Figure 1 shows the results of the analysis, with the recovered parameter estimates plotted against the generating parameters. In each panel, wider error bars indicate greater uncertainty and accuracy is shown by dispersion from the centre line. To characterize the posterior mean parameter estimates, we also present mean absolute bias error (MABE = $\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left|\widehat{\theta_i} - \theta_i\right|$) and normalized root mean square error ($NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta_i - \widehat{\theta_i}\right)^2}}{\max\theta - min(\theta)} \cdot 100$), where $\hat{\theta}$ is the estimated posterior mean parameter value and θ is the true parameter value. NRMSE provides a descriptive summary of the apparent scaled relative differences in error across parameters.

Figure 1. Recovery of group-estimated unified model parameters from an applied data structure.



Note: The *x*-axis indexes the generating parameter values. The black circles correspond to estimated posterior means of the parameters and the grey lines correspond to 95% credible intervals. The line $y = x$ is depicted, with estimates near the line indicating good recovery. Parameter ranges roughly correspond to the number of lapses on a 10-minute PVT. Parameter were bounded by transforming the homeostatic parameters *L (0)* and *S (0)*: *l (0)* = *L (0)* / *U* and *s (0)* = [*S (0)* - *L (0)*] / [*U* – *L (0)*], and it is these scaleless parameters that are depicted. For all other parameters, scale is specified in the panel headings.

The results show high estimation accuracy and certainty was present for three critical parameters: *U* which informs the relative upper-bound contribution of the homeostatic process (MAB = 0.28, NRMSE = 2.38%); κ which controls the relative contribution of the circadian process (MAB = -0.14, NRMSE = 4.2%); and Φ which controls circadian phase (MAB = -0.18, NRMSE = 2.01%). The time constant parameters that control the rate of fatigue accumulation τ_w (MAB = 0.04, NRMSE = 8.08%) and recovery τ_s (MAB = 0.25, NRMSE = 5.71%) were mostly accurate, but estimation was highly uncertain. The parameter τ_{LA} controlling long-term sleep deprivation processes recovered particularly poorly (MAB = 5.85, NRMSE = 27.99%). One Bayesian technique to address this would be to place a tight prior distribution on the value of τ_{LA}, centred on the parameter values obtained from previous studies. The recovery of the initial level of homeostatic fatigue *(S0*; MAB = -0.14, NRMSE = 21.83%*)* was also poor, but this is not necessarily problematic as the initial level of fatigue is unlikely to have a long-running effect (particularly over extended timeframes).

The analysis presented here is probably near the upper limit on expected parameter estimation in field settings. We included a large sample with a high within-person measurement sampling rate over a broad time scale. The fatigue observations here were generated assuming that the unified model is the true model of fatigue dynamics, and assuming normally distributed noise without any systematic biases. In other words, our analysis does not assume there are any additional factors (e.g., workload or fatigue countermeasures) that bear systematic influence on fatigue, which is not what would be expected in naturalistic environments. We have also assumed individuals within the sample are homogenous in terms of parameters (e.g., identical circadian phase), and we did not place constraining bounds on the data ranges that the models can predict (e.g., by fixing the minimum or maximum number of lapses). In realistic field conditions, these assumptions are likely to be violated (i.e., differences in parameters across individuals and bounded possible observed fatigue scores), reducing the quality of estimation.

Overall, the results here are consistent with prior research on the unified model (Liu et al., 2017). The relative contributions of the homeostatic and circadian process to fatigue recovered reasonably well. Although circadian phase also recovered quite well, in practice it would be more appropriate to estimate phase using alternative data (e.g., core body temperature, sleep timing, light exposure) (Brown et al., 2021; Stone et al., 2020). The parameters requiring most attention were the time constants of the homeostatic process. These parameters recovered poorly, implying that the time course of the fatigue response to sleep/wake time was difficult to estimate, even under ideal conditions. As identifying the time course of the homeostatic process is one of the primary interests of field estimation, this result suggests there may be little added utility provided by BMMs estimated from field data as compared with standard BMMs trained on laboratory data. However, there are other potential mechanisms of model advancement which we explore in more detail below, including model individualization and extension.

2.2 Model Individualization and Real-Time Prediction

The so called "Holy Grail in fatigue and performance modeling" is complete fatigue prediction individualization through tailoring model parameterizations to the individual person (Reifman, 2004, p. A177). The theory of individualization is that between-person differences in circadian phase, or potentially the biological dynamics governing sleep regulation, can be directly included within the modelling framework by adjusting parameterizations for each person uniquely. Consistent with group-level estimation, parameters can be estimated based on either observation within a controlled laboratory context, or from data collected in field operations. This latter approach is the basis of "real-time" fatigue-prediction tools, in which parameters are estimated for individuals in response to real-time incoming data streams, with the possibility of reactively identifying fatigue risks in the workplace (e.g., see Liu et al., 2017). Although this approach requires considerable amounts of data per person (i.e., both sleep and fatigue observations), there are strong justifications to assume that characterizing between-person differences may improve model performance.

From a practical perspective, in many safety-critical workplaces it is often of most interest to obtain fatigue projections for specific employees over a period of hours to a few days (fitness to work), rather than whether on average they stay below fatigue safety thresholds given a particular roster. In complex, inter-dependent work systems, such as those often required in extreme work environments, unsafe levels of fatigue in even one team member could have serious consequences (Cham et al., 2021). The real-time fatigue forecasts made possible by individualization promise improved tactical decision making (e.g., deciding on ideal times to execute mission scenarios) and crew rotation decisions (e.g., which staff may be at heightened performance risk).

There are also strong justifications from research and theoretical perspectives. Individuals are known to vary with respect to chronotype (Brown et al., 2021), the timing of rest-periods (Archer et al., 2008), and vulnerability to sleep deprivation (Chua et al., 2019). Early research has indicated individualized models are possible (Dawson et al., 2011; Ramakrishnan et al., 2015) with uncertainty regarding which parameters should be considered as stable trait difference, relative to state differences that may fluctuate within-person (Ramakrishnan et al., 2015; Van Dongen et al., 2007). A key benefit of individualization is that BMM parameters can be informed by measures other the performance criterion. For instance, recent circadian modelling research has indicated that lighting conditions bear strong predictive influence over circadian angle of entrainment and preferred sleep timing (Papatsimpa et al., 2021; Phillips et al., 2019). Allowing BMM circadian phase parameters to be informed from actigraphy and photometry data passively (see Brown et al., 2021) would reduce model complexity and improve predictive accuracy. Thus, leaving individualized parameter estimation using behavioral data for only a subset of the full model parameters.

Despite the appeal of BMM individualization, there are several notable challenges in implementing the approach. The parameter recovery issues outlined in the last section apply even more strongly when the requirement is to tune BMM parameters to the sparse fatigue observations of each individual. As a result, existing individualized models either require prior knowledge of a reliable group-average model (e.g., Liu et al., 2017; Van Dongen et al., 2007), or are applied to unrealistically simplified conditions such as total sleep deprivation

(Rajaraman et al., 2008, 2009; Van Dongen et al., 2007). Further, in field contexts there are no guarantees that the variation in fatigue observations are uniquely associated with the processes assumed within the BMM (unlike laboratory contexts whereas many factors as possible are controlled for) (Reifman et al., 2007). For example, if employees face high levels of work-induced fatigue during waking hours, and this is not directly instantiated in the BMM, it is likely that BMM fitting would falsely attribute this work-induced fatigue to increased homeostatic pressure. Further, the extent of inter-individual differences in vulnerability to sleep loss can depend on the performance measure (see Chua et al., 2019). Thus, for operational contexts, the relationship between an individual's actual task performance and model prediction may depend on the variable used in the model.

In summary, individualizing BMMs is a priority of next-generation modelling and promises many potential benefits. Although fitting BMMs to the behavioral data of individuals holds some promise towards this goal, it is constrained by substantial data limitations. Future approaches to individualization, particularly involving field data, are likely to greatly benefit from incorporating other individualized sources of data, such as light exposure (Phillips et al., 2019; Stone et al., 2020)

2.3 Extending the BMM Processes

In real-world conditions, the causes of fatigue are heterogenous and are not driven purely by homeostatic and circadian processes (S. Banks et al., 2019; Desmond & Hancock, 2001; Wilson et al., 2021). To accurately model these exogenous influences, and thereby increase prediction accuracy, parametric model extension has been pursued as a key direction for future BMMs. This involves adjusting model equations to directly specify how additional processes of interest affect the functional form of fatigue. Conventional BMMs predict fatigue based purely on sleep history and time of day, with some including processes for chronic sleep restriction (e.g., Rajdev et al., 2013). Research has focused predominantly on the influence of pharmaceutical fatigue counter-measures (Ramakrishnan et al., 2013), chronic sleep dept (Rajdev et al., 2013), and task demands or workload (Honn et al., 2016; H. T. Peng et al., 2018).

Here we consider whether parametric model extension of BMMs is likely to meet next-generation demands. To illustrate parametric model extension, and assumptions or decisions embedded within, we formally detail a worked example of the extension process. We focus on the salient example of how workload may modulate fatigue in order to demonstrate the benefits and barriers involved in parametric model extension. It is uncontroversial that fatigue can be influenced by work factors, such as shift duration and workload (Desmond & Hancock, 2001; Dorrian et al., 2011; Grech et al., 2009; Wilson et al., 2021). Consequently, there has been much discussion of extending BMMs to describes the process of how work demands (or simply work hours) influence fatigue.

Consistent with the recovery analysis, we selected the unified model of performance as the starting point for the workload extension (Rajdev et al., 2013). Our extended model includes an additional process wherein fatigue from work demands, referred to as *D*, accrues over time spent working, with the exact rate dependent upon the level of homeostatic fatigue (i.e., fatigue resulting from sleep processes). The model specifies that work demands primarily influence an individual's *sensitivity* to fatigue (Baulk et al., 2007). That is, task demands only additively increase fatigue when homeostatic pressure is high. The model implicates that high work demands can be more effectively managed by well-rested individuals with lower performance costs, relative to individuals with high homeostatic fatigue. This is consistent with how other groups have implemented workload BMM extensions. For example, Peng et al. (2018) proposed a model in which working-related fatigue accrues over time spent working, at a rate proportional to workload and current homeostatic fatigue. Further, Honn et al. (2016) incorporated a similar approach into the McCauley state-space model.

Equation 1 specifies how workload-related fatigue accrues over time spent working, with the speed of accrual at each time depending on homeostatic pressure. Equation 2 specifies how recovery from work-related fatigue follows an exponential function. A computational implementation of the model can be found in the supplementary materials.

$$D_t = D_0 + \gamma \times \int_0^t \max(S_t, 0)\, dt \tag{1}$$

$$D_t = D_l - e^{-\frac{t}{\tau_r}}(D_l - D_0) \tag{2}$$

In both equations, t represents the total time spent working or resting, D_t represents the total fatigue from work demands at *t* hours, D_0 represents the initial level of work-related fatigue (at the start of a rest or work episode), γ is a free parameter that controls the rate of fatigue accrual due to work demands, S_t represents the homeostatic pressure after working for time t. The integral of S_t is only taken for values above 0, to avoid the possibility of negative work-induced fatigue (i.e., work decreasing fatigue). Finally, τ_r is a time constant controlling the rate of recovery. Figure 6 shows the impact of adding this workload-related fatigue module onto the overall fatigue predictions from the unified model for the standard 9-5 working arrangement (across 16 different *r* parameter values).

Figure 6. Sensitivity plot for work fatigue process on total fatigue in the simulation.

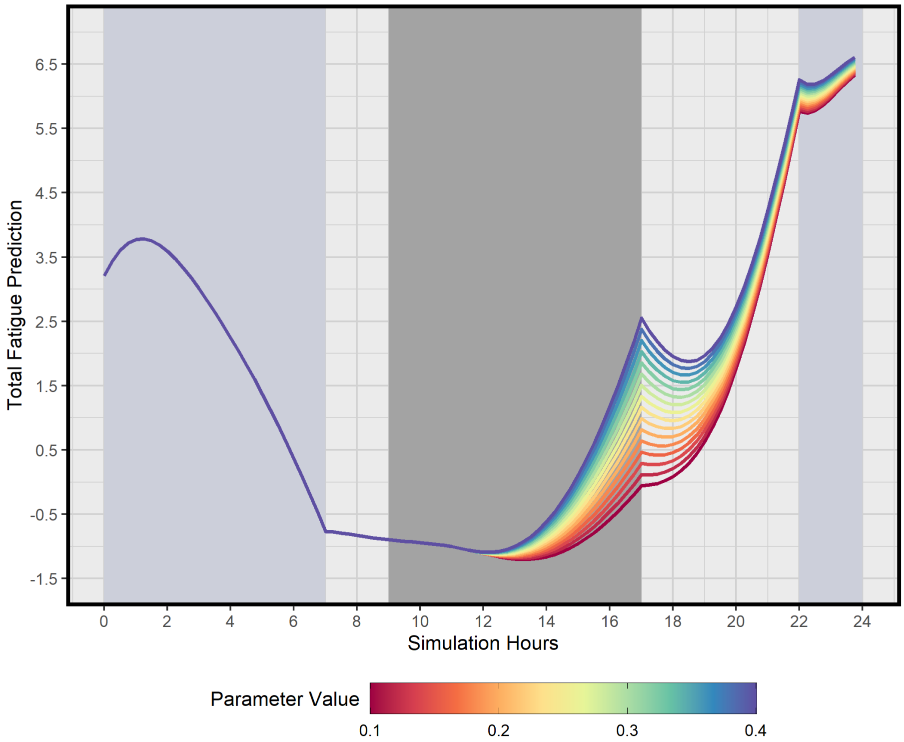


Figure note. The grey area denotes time at work (9AM-5PM), while the pale-blue area denotes sleep. The y-axis shows the total level of fatigue accrued during the simulation, while the x-axis shows the time into the simulation (in decimal hours). The parameter modulated is the work-related fatigue accrual rate, γ. In the plot, 16 values of this beta coefficient are modulated (range from 0.1 to 0.4, incremented by .02). Prior to the first work episode (dark grey), all participants have an identical level of fatigue. Note the increase in fatigue during second sleep is due to the strong circadian influence in this simulation.

The workload model above is representative of the typical process of a parametric BMM extension. Evidently, extensions require researchers to make explicit assumptions about the functional form of fatigue accrual and recovery processes, and how the added process links to the criterion performance variable. This is conceptually straightforward to implement, and in our experience, simple sensitivity estimates are useful for practitioners to identify possible high-risk situations. However, the example above also highlights challenges that prevent this approach from solving the question of next-generation predictive performance gains.

The obvious limitation of this model, shared with many other BMM extensions, is the lack of comprehensive validation. Ideally, researchers progress beyond a proposed model extension towards a well validated model component with theoretical rigor (e.g., the sleep-inertia component of the three-process model, Åkerstedt & Folkard, 1997). Validation requires many of the same considerations as those for parameter estimation discussed in section 2.1. For example, to determine appropriate parameterizations, researchers would need to conduct a controlled laboratory study in which workload was systematically manipulated. It must also be determined if work-related fatigue processes elicited using simple laboratory tasks are adequate for generalizing to situated workplace contexts. Honn et al. (2016) developed a workload extension using PVT performance of pilots performing simulated take-offs and landings. Their workload model was calibrated by estimating a parameter φ, that controlled how severely cognitive task load impacted fatigue. All other model parameters were fixed, presumably to enable estimation. Such tightly constrained approaches are useful during initial development, but neglects possible parameter trade-offs, raising concerns of model identifiability. Indeed, the recovery behavior of even baseline BMMs (section 2.1.2) suggests significant challenges exist in freely estimating extended BMMs.

Arguably, the most significant challenge associated with parametric model extension is grappling with the complexity growth of model exploration and validation. Unlike more descriptive conventional statistical approaches (e.g., general linear modelling), BMMs attempt to precisely specify the functional form of their component processes, and the relationships of the processes and their relationships to each other. The current state of knowledge of complex forms of fatigue, such as work-induced fatigue, provide few constraints on the most appropriate model form. Ideally, model extensions should be compared to theoretical viable alternative models. For instance, the workload model we introduced above could account for situations of underload induced fatigue (Shultz et al., 2010; Young & Stanton, 2002) and does not account for the known “carry over” effects of high workload situations on subsequent sleep (Crain et al., 2018). Each such point of complexity needs to be weighed against the relative improvement in predictive gains offered. Further, such model selection is likely to suffer from identifiability issues analogous to the parameter identifiability issues discussed earlier, given the numerous theoretically plausible ways that work could affect fatigue, and the relative scarcity of work and fatigue data.

Despite these critiques, parametric BMM extensions do hold clear benefits. Practically, even approximate estimations of how factors such as pharmaceutical counter-measures impact fatigue can inform the development of safety-promotion strategies (Reifman et al., 2016, 2019). Similarly, imperfect models of workload still provide a methodology for practitioners to evaluate possible high-risk roster scenarios and formalize assumptions in a manner that would otherwise remain as qualitative verbal theory (Ballard et al., 2021). These techniques thus offer practical benefits to practitioners, but are unlikely to provide step-changes in predictive accuracy, theoretical advancement, or operational safety gains.

2.4 Joint and Non-Parametric Modelling

The key aim of next-generation fatigue models is to enhance our prediction in ways more relevant to the individual operator and the context in which they are situated. This can involve both increased precision in multi-factorial prediction as well as improving our knowledge of the theoretically relevant factors underpinning fatigue. Given the limitations of BMMs, there is a need to examine the alternative methodological approaches, and how they may help address these goals. A promising direction frequently noted is adoption of non-parametric, or machine-learning, based approaches.

Reifman (2004) distinguished between parametric fatigue models (i.e., BMMs), and non-parametric fatigue models (i.e., machine learning approaches), such as artificial neural networks, which can generate predictions without necessarily requiring an *a priori* model structure (see also Breiman, 2001). A core benefit of machine learning approaches is that, in principle, they can incorporate any number of predictors (Jordan & Mitchell, 2015), including urine output, cortisol levels, workload, and light exposure (Reifman, 2004). Reifman (2007) proposed several variants of fatigue prediction involving non-parametric approaches, including so-called "hybrid methods" in which BMMs could even be embedded within neural network architectures to support prediction. Machine learning may also better enable integration of real-time physiological indicators of fatigue, such as cardiovascular state (Aryal et al., 2017; Hu & Lodewijks, 2020), although to-date these have been argued to have limited utility and validity (Dawson et al., 2014). Given sufficiently mature data and computational infrastructure, machine learning approaches could predict fatigue during operations, potentially recommending interventions when fatigue is likely to be high. The prospect of deploying such systems, in even extreme workplace environments, is increasingly possible due to advances in statistical methods, data storage systems, and computational processing power.

It would be timely to begin identifying and validating applications of machine learning methods (and related nonparametric approaches) into the fatigue prediction toolkit. Even in the case of researchers who continue to pursue BMM extension, machine learning methods may also be of significant advantage by integrating with BMMs indirectly. For instance, by generating estimates of individuals' sleep quality and quantity from wearable technologies — referred to as *activity classification* (Lewicke et al., 2008; Piotrowski & Szypulska, 2017). Indeed, Sundararajan et al. (2021) recently applied used random forest machine learning models to classify wrist-worn accelerometry data into sleep/wake and non-wear. The approach was superior to existing methods and is accessible under a direct access license. BMMs and machine learning could function synergistically, for example by using BMMs to improve machine learning predictions (e.g., model fusion), or by using nonparametric approaches to model whatever residual performance data cannot be fitted by a BMM (Bourgin et al., 2019; Sense et al., 2021). This may help overcome some of the limitations raised with estimating BMMs directly.

Due to their lack of a priori structure, machine learning methods have substantial data requirements, and can only provide 'black box' fatigue estimates that are difficult to decompose into the underlying phenomenological processes (e.g., circadian rhythm, homeostatic). The lack of a specified underlying process can lead to unexpected and intractable failures when predicting data outside of the model's range of training (e.g., when simulating alternative sleep schedules or attempting to generalize predictions across different work environments). Recently, Cochrane et al., (2021) used an ensemble machine learning model to predict the performance effects of sleep-loss in a forced desynchrony protocol dataset. However, the data requirements for accurate prediction were significant, requiring required 10-minute PVT administration every 2-8 hours, and is still grounded in laboratory validation. Nevertheless, there is a range of emerging non-parametric methods that may support fatigue researchers in development predictive frameworks that exceed the capabilities afforded by parametric BMMs alone. Overall, this would appear to be the most promising pathway forward for the next-generation of BMM research and practice.

3 Summary, Future Directions and Conclusions

The purpose of this paper was to examine whether BMMs are sufficiently able to meet the demands of prominent next-generation modelling requirements and to identify limits of current methods and avenues for future research. We outlined themes from the literature regarding what is needed for improved models in operational contexts, dating back over 15 years. We described and evaluated current directions for advancing the next-generation of fatigue prediction methods, focusing on the application of BMMs in operational contexts.

Firstly, we evaluated the practice of tailoring model parameters to populations or individuals of interest using in-situ data. Despite the conceptual appeal, we found no strong evidence from the literature supporting the feasibility of this approach and noted the logistical challenges to data collection are high. We then conducted a parameter recovery study, focused on a naval submarine context, which revealed that even under optimistic modelling conditions, for several critical parameters the estimation using field data was likely to produce highly uncertain estimates, at least for particular parameters. This finding lends support to Dawson's (2011) suggestion that it may not be possible to tailor BMMs to provide accurate forecasts for a workplace context using field data. Secondly, we reviewed the work to-date regarding individualization. While we noted some promising advancements have been made in laboratory contexts, overall, individual estimation in such work was constrained to tightly controlled parameterizations of a subset of parameters. Thirdly, we examined the practice of parametric BMM extension, and focused on the case of extending a BMM to incorporate the effects of workload. Our example and review highlighted the substantial challenges associated with extending BMMs, noting that BMMs require tightly constrained and theoretically informed mappings between moderator variables and both the underlying fatigue function (e.g., sleepiness) and the interaction with other variables of interest. We concluded that next-generation BMMs involving model extensions are likely to be prohibitively difficult to accurately specify, and extremely challenging to validate complex model processes.

It is crucial to emphasize that the limitations and barriers we have reviewed do not preclude continued use of BMMs for their intended purpose of risk-mitigation in average-level scheduling. Moreover, we are not suggesting there is no merit in continuing to pursue existing goals for next-generation BMM features such as generating fatigue predictions that are targeted to populations, individuals, or that can incorporate domain-specific variables such as workload. However, our review highlights significant barriers to achieving these goals with BMMs. In light of these barriers, in the following section, we provide a set of recommendations that can guide researchers and model developers toward what we believe are more fruitful avenues for advancing fatigue science.

3.1 Recommendations for Advancing Fatigue Modelling

3.1.1 Expanding Theoretical Frameworks of Fatigue and Performance

The parametric equations underlying the BMM approach are inherently only able to capture some proportion of the rich and complex dynamics underlying fatigue. Moving forward, fatigue science will require increased emphasis on multi-model thinking, both numerically and conceptually, and a change in how we approach measurement of fatigue itself. An important pathway to improve fatigue prediction is to better specify the relationship between fatigue and the performance predictions of interest. Our recovery analysis revealed that the behavioral response to fatigue is the element that BMMs are least effective in capturing (i.e., homeostatic process). There are several possible avenues to this end emerging in current literature.

A particularly important shift is the need to move away from relying on coarse data (e.g., PVT mean response time, lapse rates, and subjective fatigue ratings) for both model validation and theoretical innovation. Such reliance restricts BMMs to only output predictions in terms of those measures, limiting their relevance to performance of complex workplace tasks (Williamson et al., 2011). Validation efforts to date for commercial models has only linked coarse predictions based on sleep opportunity against historical safety incidents (Hursh et al., 2006). Others have suggested that models be calibrated against task performance metrics obtained from real-world scenarios, or representative simulations (Reifman et al., 2007). This would improve ecological validity (for the relevant domain) but do little for generalization. The alternative pathway is to better model the latent mechanisms underlying performance to ensure that performance predictions can generalize across a wide range of scenarios and contexts.

3.1.1.1 Computational Cognitive Models

To better understand how behavioral performance changes as a function of fatigue, we need to develop or adopt dynamic models of behavior. There have been significant developments in computational cognitive models that specify the processes that underlie task performance in detail, and thus provide an appropriate means to quantify the effects of fatigue on performance. For example, PVT metrics such as number of lapses and mean RT have ambiguous mappings to underlying cognitive processes (Chua et al., 2019; Veksler & Gunzelmann, 2018). It is unclear whether fatigue increases mean RT and number of lapses because individuals process information less efficiently when fatigued, or because they require more evidence to respond (i.e., increase caution). These competing explanations have direct implications for the safety profile of tasks under fatigue states, but they can be compared by applying evidence accumulation models that use detailed response choice and response time data to measure underlying cognitive constructs such as processing speed and caution. Indeed, evidence accumulation models have been applied PVT performance (Chavali et al., 2017; Ratcliff & Van Dongen, 2011), and integrated with BMMs to a limited extent (Walsh et al., 2017). These early findings implicate fatigue being associated with processing speed deficits in the PVT rather than response caution. Alternative work has incorporated fatigue into the ACT-R cognitive architecture (Baradaran Khosroshahi, 2019; Gunzelmann et al., 2019) which can model a broad range of cognitive tasks, offering opportunities to generalize fatigue-related performance predictions (Gunzelmann et al., 2019).

Sophisticated computational models of behavior help to better understand the differential causes of the fatigue response. The basic forms of all BMMs, including the workload extensions, generally ground fatigue accumulation as largely a homeostatic driven process. This research is founded on findings with the PVT, initially selected as it is sensitive sleep-loss. The PVT is also sensitive to many other biases which BMMs do not directly examine. For example, Hockey (2013) argue that in many circumstances, fatigue can be manifest as a motivational issue with more transient influences on performance.

3.1.1.2 Dynamic Longitudinal Models

In considering alternative methods for understanding the latent dynamics underlying fatigue, a crucial relevant development, particularly for when considering field measurement, is the application of models that how fatigue interacts and coevolves with other individual, environmental, and work-related factors over time. For instance, dynamic structural equation modelling (Asparouhov et al., 2018) and the continuous time counterpart (Driver & Voelkle, 2018; Ryan et al., 2018) allows researchers to model the dynamics of observed and latent variables. That is, how variables evolve and relate to each other over time (i.e., auto- and cross-regressive effects). These dynamic modelling approaches may provide insights into questions of causality between measures, and unlike BMMs do not require the exact mathematical specifications of interactions between factors. Dynamic structural equation models therefore can potentially detect longer-term ‘knock on’ effects of workload to sleep quality and quantity, and long-term burnout (Crain et al., 2018; Wilson et al., 2021). It is also important to consider that many of the variables related to fatigue are state dependent, that is the causal relationships among variables change with different states of the system (Chang et al., 2017; Sugihara & May, 1990). For instance, a sustained level of high workload may cause fatigue, but high levels of fatigue may reduce perceived cognitive capacity causing higher perceptions of workload (e.g., on longer-time scales, Guthier et al., 2020). Emerging methods such as *empirical dynamic modeling* (Chang et al., 2017) offer a means to decompose such complex interdependencies in causal systems. Adoption of these methods can improve our ability to design better work, mitigate risk, and support individuals working in demanding and extreme environments.

3.1.2 Improve Open Science Practices

Open science practices are increasingly perceived as integral to the conduct of robust science. To accelerate research into fatigue prediction, we propose the field should place significant efforts to offering greater computational reproducibility and transparency. Presently, most fatigue prediction solutions are closed source and proprietary, and in many cases, this can make independent replication, comparison to, or extension of the work reported in scientific articles impractical or impossible. Generally, in cases where BMM formulae are provided, the respective computational implementations are not. There are strong arguments for going beyond this minimum state of reproducibility of providing only formulae, towards a gold standard in which flexible model implementations are provided with journal articles (G. C. Banks et al., 2018; R. D. Peng, 2011; Wilson et al., 2019). Notably, gold standard reproducibility lowers the 'cost to entry' for researchers to adopt fatigue prediction methods, inviting researchers with unique perspectives and skill sets that may be well suited to advance the other frontiers outlined in this article. It also assures that model implementation details (e.g., regarding parameter estimation) can be understood by outsiders, which is difficult to assure from textual descriptions alone. We note that initial steps towards computational reproducibility have been taken with the development of *2B-Alert Web*, an open-access web application that allows users to graphically examine the predictions of a BMM (Reifman et al., 2016), and the release of an open-source R package for BMMs, *FIPS* (Wilson et al., 2020). FIPS provides sleep and fatigue data structures, BMM implementations, and utility functions. Efforts such as these are crucial to encouraging cumulative science, mitigating research fragmentation, and preventing duplicated efforts.

Another valuable open scientific practice is data sharing (for an introduction to ethical data sharing practices, consult Meyer, 2018). Publicly available sleep and fatigue data is currently extremely hard to come by, for both datasets of field and laboratory contexts. The significant benefits and challenges of open-data practices have been widely discussed (G. C. Banks et al., 2018; Gewin, 2016) and there is clear consensus that engaging such practices accelerates methodological innovation and scientific discovery. We propose that fatigue prediction researchers engage substantially more in open data sharing.

In the field of fatigue prediction, open data sharing would enable newly developed fatigue prediction methods to be evaluated on legacy ‘benchmark’ datasets. This would avoid the common situation where BMMs are extended to account for new datasets and contexts, but it is unclear whether the extended BMMs still adequately fit the data that the original models were developed on (Reifman & Gander, 2004). Further, it would enable computational modelling experts lacking the resources of a sleep laboratory to advance fatigue prediction methods and create opportunities for pooling field and laboratory-based data. These advances would translate to practical benefits for the employees and organizations that these models are intended to support. Indeed, the Flight Safety Foundation (2005) review into BMMs indicated a need for data sharing practices, and the importance of these practices in building trust with industry practitioners.

We recognize that individuals can face barriers to open science practices, including the ethical concerns regarding data privacy, intellectual property and stakeholder concerns, and the need to justify the substantial investments in initially obtaining relevant data. We do not wish to discount these concerns or claim to resolve them. Indeed, practical concerns such as funding have long posed difficulties for the advancement of fatigue prediction methods (Akerstedt et al., 2004). However, such challenges do not undermine the prospective value of the practices we have raised. Ultimately, engagement in open science practices is required if we are to overcome the barriers to achieve next-generation model features.

3.2 Conclusions

BMMs have played an instrumental role in the implementation of evidence-based fatigue management strategies in safety-critical contexts. There remains significant interest in the development of next-generation BMMs capable of providing tailored and more accurate fatigue predictions. This paper has provided review and analysis of several of the key directions underpinning the development of next-generation models, and has revealed there are significant challenges to the realization of benefits from conventionally proposed advancements. Just as fatigue management strategies require consideration of multiple factors, fatigue prediction methods appear to require the implementation of multi-model approaches. The integration and fusion of BMMs with other models, including approaches such as cognitive modelling and machine learning, will be most critical to support more targeted, relevant, and accurate fatigue prediction in safety-critical workplaces.

References

Åkerstedt, T., & Folkard, S. (1997). The Three-Process Model of Alertness and Its Extension to Performance, Sleep Latency, and Sleep Length. *Chronobiology International*, *14*(2), 115–123. https://doi.org/10.3109/07420529709001149

Akerstedt, T., Folkard, S., & Portin, C. (2004). Predictions from the three-process model of alertness. *Aviation, Space, and Environmental Medicine*, *75*(3 Suppl), A75-83. https://doi.org/10.3109/07420529709001149

Archer, S. N., Viola, A. U., Kyriakopoulou, V., von Schantz, M., & Dijk, D.-J. (2008). Inter-Individual Differences In Habitual Sleep Timing and Entrained Phase of Endogenous Circadian Rhythms of BMAL1, PER2 and PER3 mRNA in Human Leukocytes. *Sleep*, *31*(5), 608–617.

Aryal, A., Ghahramani, A., & Becerik-Gerber, B. (2017). Monitoring fatigue in construction workers using physiological measurements. *Automation in Construction*, *82*, 154–165. https://doi.org/10.1016/j.autcon.2017.03.003

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Ballard, T., Palada, H., Griffin, M., & Neal, A. (2021). An Integrated Approach to Testing Dynamic, Multilevel Theory: Using Computational Models to Connect Theory, Model, and Data. *Organizational Research Methods*, *24*(2), 251–284. https://doi.org/10.1177/1094428119881209

Banks, G. C., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2018). Answers to 18 Questions About Open Science Practices. *Journal of Business and Psychology*. https://doi.org/10.1007/s10869-018-9547-8

Banks, S., Landon, L. B., Dorrian, J., Waggoner, L. B., Centofanti, S. A., Roma, P. G., & Van Dongen, H. P. A. (2019). Effects of fatigue on teams and their role in 24/7 operations. *Sleep Medicine Reviews*, *48*, 101216. https://doi.org/10.1016/j.smrv.2019.101216

Baradaran Khosroshahi, E. (2019). *A Unified Model of Fatigue in a Cognitive Architecture* [Ph.D., Drexel University]. https://search.proquest.com/docview/2328377233/abstract/C3799B09924944F1PQ/1

Beare, A. N., Bondi, K. R., Biersner, R. J., & Naitoh, P. (1981). Work and rest on nuclear submarines. *Ergonomics*, *24*(8), 593–610. https://doi.org/10.1080/00140138108924882

Bell, S. T., Fisher, D. M., Brown, S. G., & Mann, K. E. (2016). An Approach for Conducting Actionable Research With Extreme Teams: *Journal of Management*. https://doi.org/10.1177/0149206316653805

Bendak, S., & Rashid, H. S. J. (2020). Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics*, *76*, 102928. https://doi.org/10.1016/j.ergon.2020.102928

Bishop, S. L. (2004). Evaluating teams in extreme environments: From issues to answers. *Aviation Space and Environmental Medicine*, *75*(7 SEC. 1), 14–21.

Boeing, A. A., Jorristma, K., Griffin, M. A., & Parker, S. K. (2020). Surfacing the social factors early: A sociotechnical approach to the design of a future submarine. *Australian Journal of Management*, *45*(3), 527–545. https://doi.org/10.1177/0312896220920338

Bourgin, D. D., Peterson, J. C., Reichman, D., Russell, S. J., & Griffiths, T. L. (2019). Cognitive model priors for predicting human decisions. *International Conference on Machine Learning*, 5133–5141. http://proceedings.mlr.press/v97/peterson19a.html

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–215.

Brown, L. S., Hilaire, M. A. S., McHill, A. W., Phillips, A. J. K., Barger, L. K., Sano, A., Czeisler, C. A., Doyle, F. J., & Klerman, E. B. (2021). A classification approach to estimating human circadian phase under circadian alignment from actigraphy and photometry data. *Journal of Pineal Research*, *71*(1), e12745. https://doi.org/10.1111/jpi.12745

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software, Articles*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Cham, B. S., Boeing, A. A., Wilson, M. D., Griffin, M. A., & Jorritsma, K. (2021). Endurance in extreme work environments. *Organizational Psychology Review*, 20413866211006440. https://doi.org/10.1177/20413866211006441

Chang, C.-W., Ushio, M., & Hsieh, C. (2017). Empirical dynamic modeling for beginners. *Ecological Research*, *32*(6), 785–796. https://doi.org/10.1007/s11284-017-1469-9

Chavali, V. P., Riedy, S. M., & Van Dongen, H. P. A. (2017). Signal-to-Noise Ratio in PVT Performance as a Cognitive Measure of the Effect of Sleep Deprivation on the Fidelity of Information Processing. *Sleep*, *40*(3), zsx016. https://doi.org/10.1093/sleep/zsx016

Chua, E. C.-P., Sullivan, J. P., Duffy, J. F., Klerman, E. B., Lockley, S. W., Kristal, B. S., Czeisler, C. A., & Gooley, J. J. (2019). Classifying attentional vulnerability to total sleep deprivation using baseline features of Psychomotor Vigilance Test performance. *Scientific Reports*, *9*(1), 12102. https://doi.org/10.1038/s41598-019-48280-4

Civil Aviation Safety Authority. (2014). *Biomathematical Fatigue Models Guidance Document*.

Cochrane, C., Ba, D., Klerman, E. B., & St. Hilaire, M. A. (2021). An ensemble mixed effects model of sleep loss and performance. *Journal of Theoretical Biology*, *509*, 110497. https://doi.org/10.1016/j.jtbi.2020.110497

Crain, T. L., Brossoit, R. M., & Fisher, G. G. (2018). Work, Nonwork, and Sleep (WNS): A Review and Conceptual Framework. *Journal of Business and Psychology*, *33*(6), 675–697. https://doi.org/10.1007/s10869-017-9521-x

Dawson, D. (2012). Fatigue research in 2011: From the bench to practice. *Accident Analysis & Prevention*, *45*, 1–5. https://doi.org/10.1016/j.aap.2011.09.016

Dawson, D., Darwent, D., & Roach, G. D. (2017). How should a bio-mathematical model be used within a fatigue risk management system to determine whether or not a working time arrangement is safe? *Accident Analysis & Prevention*, *99*, 469–473. https://doi.org/10.1016/j.aap.2015.11.032

Dawson, D., Ian Noy, Y., Härmä, M., Åkerstedt, T., & Belenky, G. (2011). Modelling fatigue and the use of fatigue models in work settings. *Accident Analysis & Prevention*, *43*(2), 549–564. https://doi.org/10.1016/j.aap.2009.12.030

Dawson, D., Searle, A. K., & Paterson, J. L. (2014). Look before you (s)leep: Evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. *Sleep Medicine Reviews*, *18*(2), 141–152. https://doi.org/10.1016/j.smrv.2013.03.003

Dean, D. A., Fletcher, A., Hursh, S. R., & Klerman, E. B. (2007). Developing Mathematical Models of Neurobehavioral Performance for the "Real World." *Journal of Biological Rhythms*, *22*(3), 246–258. https://doi.org/10.1177/0748730407301376

Desmond, P. A., & Hancock, P. A. (2001). Active and passive fatigue states. In *Stress, workload, and fatigue.* (pp. 455–465). Lawrence Erlbaum Associates Publishers.

Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*, *23*(4), 774–799. https://doi.org/10.1037/met0000168

Flight Safety Foundation. (2005). *The Singapore Experience: Task Force Studies Scientific Data to Assess Flights*. *24*(8-9 (August–September 2005)), 64.

Flynn-Evans, E. E., Kirkley, C., Young, M., Bathurst, N., Gregory, K., Vogelpohl, V., End, A., Hillenius, S., Pecena, Y., & Marquez, J. J. (2020). Changes in performance and bio-mathematical model performance predictions during 45 days of sleep restriction in a simulated space mission. *Scientific Reports*, *10*(1), 15594. https://doi.org/10.1038/s41598-020-71929-4

Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A., & Popkin, S. (2011). Fatigue risk management: Organizational factors at the regulatory and industry/company level. *Accident Analysis & Prevention*, *43*(2), 573–590. https://doi.org/10.1016/j.aap.2009.11.007

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature*, *529*(7584), 117–119. https://doi.org/10.1038/nj7584-117a

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*(2), 273–286. https://doi.org/10.1016/j.chb.2004.02.014

Gunzelmann, G., M. James, S., & Caldwell, J. L. (2019). Basic and applied science interactions in fatigue understanding and risk mitigation. In *Progress in Brain Research* (Vol. 246, pp. 177–204). Elsevier. https://doi.org/10.1016/bs.pbr.2019.03.022

Guo, J.-H., Ma, X.-H., Ma, H., Zhang, Y., Tian, Z.-Q., Wang, X., & Shao, Y.-C. (2020). Circadian misalignment on submarines and other non-24-h environments – from research to application. *Military Medical Research*, *7*(1), 39. https://doi.org/10.1186/s40779-020-00268-2

Guthier, C., Dormann, C., & Voelkle, M. C. (2020). Reciprocal effects between job stressors and burnout: A continuous time meta-analysis of longitudinal studies. *Psychological Bulletin*, *146*(12), 1146–1173. https://doi.org/10.1037/bul0000304

Honn, K. A., Satterfield, B. C., McCauley, P., Caldwell, J. L., & Van Dongen, H. P. A. (2016). Fatiguing effect of multiple take-offs and landings in regional airline operations. *Accident Analysis & Prevention*, *86*, 199–208. https://doi.org/10.1016/j.aap.2015.10.005

Hu, X., & Lodewijks, G. (2020). Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *Journal of Safety Research*, S0022437519306735. https://doi.org/10.1016/j.jsr.2019.12.015

Hursh, S. R., Raslear, T. G., Kaye, A. S., & Fanzone, J. F. (2006). *Validation and Calibration of a Fatigue Assessment Tool for Railroad Work Schedules, Summary Report: (736302011-001)* [Data set]. American Psychological Association. https://doi.org/10.1037/e736302011-001

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., Storm, W. F., Miller, J. C., & Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine*, *75*(3), A44–A53.

James, F. O., Waggoner, L. B., Weiss, P. M., Patterson, P. D., Higgins, J. S., Lang, E. S., & Van Dongen, H. P. A. (2018). Does Implementation of Biomathematical Models Mitigate Fatigue and Fatigue-related Risks in Emergency Medical Services

Operations? A Systematic Review. *Prehospital Emergency Care*, *22*(sup1), 69–80. https://doi.org/10.1080/10903127.2017.1384875

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Klerman, E. B., & Hilaire, M. St. (2007). Review: On Mathematical Modeling of Circadian Rhythms, Performance, and Alertness. *Journal of Biological Rhythms*, *22*(2), 91–102. https://doi.org/10.1177/0748730407299200

Lewicke, A., Sazonov, E., Corwin, M. J., Neuman, M., & Schuckers, S. (2008). Sleep Versus Wake Classification From Heart Rate Variability Using Computational Intelligence: Consideration of Rejection in Classification Models. *IEEE Transactions on Biomedical Engineering*, *55*(1), 108–118. https://doi.org/10.1109/TBME.2007.900558

Liu, J., Ramakrishnan, S., Laxminarayan, S., Balkin, T. J., & Reifman, J. (2017). Real-time individualization of the unified model of performance. *Journal of Sleep Research*, *26*(6), 820–831. https://doi.org/10.1111/jsr.12535

Mallis, M. M., & DeRoshia, C. W. (2005). Circadian rhythms, sleep, and performance in space. *Aviation, Space, and Environmental Medicine*, *76*(6 Suppl), B94-107.

Meyer, M. N. (2018). Practical Tips for Ethical Data Sharing. *Advances in Methods and Practices in Psychological Science*, *1*(1), 131–144. https://doi.org/10.1177/2515245917747656

Noy, Y. I., Horrey, W. J., Popkin, S. M., Folkard, S., Howarth, H. D., & Courtney, T. K. (2011). Future directions in fatigue and safety research. *Accident Analysis & Prevention*, *43*(2), 495–497. https://doi.org/10.1016/j.aap.2009.12.017

Papatsimpa, C., Schlangen, L. J. M., Smolders, K. C. H. J., Linnartz, J.-P. M. G., & de Kort, Y. A. W. (2021). The interindividual variability of sleep timing and circadian phase in humans is influenced by daytime and evening light conditions. *Scientific Reports*, *11*(1), 13709. https://doi.org/10.1038/s41598-021-92863-z

Peng, H. T., Bouak, F., Wang, W., Chow, R., & Vartanian, O. (2018). An improved model to predict performance under mental fatigue. *Ergonomics*, 1–16. https://doi.org/10.1080/00140139.2017.1417641

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, *334*(6060), 1226–1227. https://doi.org/10.1126/science.1213847

Phillips, A. J. K., Vidafar, P., Burns, A. C., McGlashan, E. M., Anderson, C., Rajaratnam, S. M. W., Lockley, S. W., & Cain, S. W. (2019). High sensitivity and interindividual variability in the response of the human circadian system to evening light. *Proceedings of the National Academy of Sciences*, 201901824. https://doi.org/10.1073/pnas.1901824116

Piotrowski, Z., & Szypulska, M. (2017). Classification of falling asleep states using HRV analysis. *Biocybernetics and Biomedical Engineering*, *37*(2), 290–301. https://doi.org/10.1016/j.bbe.2017.02.003

Price, J. M., & Coury, B. G. (2015). A Method for Applying Fatigue Science to Accident Investigation. *Reviews of Human Factors and Ergonomics*, *10*(1), 79–114. https://doi.org/10.1177/1557234X15573950

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rajaraman, S., Gribok, A. V., Wesensten, N. J., Balkin, T. J., & Reifman, J. (2008). Individualized performance prediction of sleep-deprived individuals with the two-process model. *Journal of Applied Physiology*, *104*(2), 459–468. https://doi.org/10.1152/japplphysiol.00877.2007

Rajaraman, S., Gribok, A. V., Wesensten, N. J., Balkin, T. J., & Reifman, J. (2009). An Improved Methodology for Individualized Performance Prediction of Sleep-Deprived Individuals with the Two-Process Model. *Sleep*, *32*(10), 1377–1392. https://doi.org/10.1093/sleep/32.10.1377

Rajdev, P., Thorsley, D., Rajaraman, S., Rupp, T. L., Wesensten, N. J., Balkin, T. J., & Reifman, J. (2013). A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation. *Journal of Theoretical Biology*, *331*, 66–77. https://doi.org/10.1016/j.jtbi.2013.04.013

Ramakrishnan, S., Lu, W., Laxminarayan, S., Wesensten, N. J., Rupp, T. L., Balkin, T. J., & Reifman, J. (2015). Can a mathematical model predict an individual's trait-like response to both total and partial sleep loss? *Journal of Sleep Research*, *24*(3), 262–269. https://doi.org/10.1111/jsr.12272

Ramakrishnan, S., Rajaraman, S., Laxminarayan, S., Wesensten, N. J., Kamimori, G. H., Balkin, T. J., & Reifman, J. (2013). A biomathematical model of the restoring effects of caffeine on cognitive performance during sleep deprivation. *Journal of Theoretical Biology*, *319*, 23–33. https://doi.org/10.1016/j.jtbi.2012.11.015

Ramakrishnan, S., Wesensten, N. J., Balkin, T. J., & Reifman, J. (2016). A Unified Model of Performance: Validation of its Predictions across Different Sleep/Wake Schedules. *Sleep*, *39*(1), 249–262. https://doi.org/10.5665/sleep.5358

Ratcliff, R., & Van Dongen, H. P. A. (2011). Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, *108*(27), 11285–11290. https://doi.org/10.1073/pnas.1100483108

Reifman, J. (2004). Alternative methods for modeling fatigue and performance. *Aviation, Space, and Environmental Medicine*, *75*(3 Suppl), A173-180.

Reifman, J., & Gander, P. (2004). Commentary on the Three-Process Model of Alertness and Broader Modeling Issues. *Aviation, Space, and Environmental Medicine*, *75*(3), 5.

Reifman, J., Kumar, K., Wesensten, N. J., Tountas, N. A., Balkin, T. J., & Ramakrishnan, S. (2016). 2B-Alert Web: An Open-Access Tool for Predicting the Effects of Sleep/Wake Schedules and Caffeine Consumption on Neurobehavioral Performance. *Sleep*, *39*(12), 2157–2159. https://doi.org/10.5665/sleep.6318

Reifman, J., Rajaraman, S., & Gribok, A. V. (2007). Moving Towards Individualized Performance Models. *Sleep*, *30*(9), 1081–1082. https://doi.org/10.1093/sleep/30.9.1081

Reifman, J., Ramakrishnan, S., Liu, J., Kapela, A., Doty, T. J., Balkin, T. J., Kumar, K., & Khitrov, M. Y. (2019). 2B-Alert App: A mobile application for real-time individualized prediction of alertness. *Journal of Sleep Research*, *28*(2), e12725. https://doi.org/10.1111/jsr.12725

Riedy, S. M., Roach, G. D., & Dawson, D. (2020). Sleep–wake behaviors exhibited by shift workers in normal operations and predicted by a biomathematical model of fatigue. *Sleep*, zsaa049. https://doi.org/10.1093/sleep/zsaa049

Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A Continuous-Time Approach to Intensive Longitudinal Data: What, Why, and How? In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous Time Modeling in the Behavioral and Related Sciences* (pp. 27–54). Springer International Publishing. https://doi.org/10.1007/978-3-319-77219-6_2

Sandal, G. M., Leon, G. R., & Palinkas, L. (2006). Human challenges in polar and space environments. *Reviews in Environmental Science and Bio/Technology*, *5*(2), 281–296. https://doi.org/10.1007/s11157-006-9000-8

Sense, F., Wood, R., Collins, M. G., Fiechter, J., Wood, A., Krusmark, M., Jastrzembski, T., & Myers, C. W. (2021). Cognition-Enhanced Machine Learning for Better Predictions with Limited Data. *Topics in Cognitive Science*, tops.12574. https://doi.org/10.1111/tops.12574

Shultz, K. S., Wang, M., & Olson, D. A. (2010). Role overload and underload in relation to occupational stress and health. *Stress and Health*, *26*(2), 99–111. https://doi.org/10.1002/smi.1268

Stone, J. E., Postnova, S., Sletten, T. L., Rajaratnam, S. M. W., & Phillips, A. J. K. (2020). Computational approaches for individual circadian phase prediction in field settings. *Current Opinion in Systems Biology*, *22*, 39–51. https://doi.org/10.1016/j.coisb.2020.07.011

Sugihara, G., & May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, *344*(6268), 734–741. https://doi.org/10.1038/344734a0

Sundararajan, K., Georgievska, S., te Lindert, B. H. W., Gehrman, P. R., Ramautar, J., Mazzotti, D. R., Sabia, S., Weedon, M. N., van Someren, E. J. W., Ridder, L., Wang, J., & van Hees, V. T. (2021). Sleep classification from wrist-worn accelerometer data using random forests. *Scientific Reports*, *11*(1), 24. https://doi.org/10.1038/s41598-020-79217-x

Van Dongen, H. P. A., Bender, A. M., & Dinges, D. F. (2012). Systematic individual differences in sleep homeostatic and circadian rhythm contributions to neurobehavioral impairment during sleep deprivation. *Accident Analysis & Prevention*, *45*, 11–16. https://doi.org/10.1016/j.aap.2011.09.018

Van Dongen, H. P. A., Mott, C. G., Huang, J.-K., Mollicone, D. J., McKenzie, F. D., & Dinges, D. F. (2007). Optimization of Biomathematical Model Predictions for

Cognitive Performance Impairment in Individuals: Accounting for Unknown Traits and Uncertain States in Homeostatic and Circadian Processes. *Sleep*, *30*(9), 1129–1143. https://doi.org/10.1093/sleep/30.9.1129

van Leeuwen, W. M. A., Pekcan, C., Barnett, M., & Kecklund, G. (2020). Mathematical modelling of sleep and sleepiness under various watch keeping schedules in the maritime industry. *Marine Policy*, 104277. https://doi.org/10.1016/j.marpol.2020.104277

Veksler, B. Z., & Gunzelmann, G. (2018). Functional Equivalence of Sleep Loss and Time on Task Effects in Sustained Attention. *Cognitive Science*, *42*(2), 600–632. https://doi.org/10.1111/cogs.12489

Vital-Lopez, F. G., Doty, T. J., & Reifman, J. (2021). Optimal sleep and work schedules to maximize alertness. *Sleep*, zsab144. https://doi.org/10.1093/sleep/zsab144

Walsh, M. M., Gunzelmann, G., & Van Dongen, H. P. A. (2017). Computational cognitive modeling of the temporal dynamics of fatigue from sleep loss. *Psychonomic Bulletin & Review*, *24*(6), 1785–1807. https://doi.org/10.3758/s13423-017-1243-6

Williamson, A., Lombardi, D. A., Folkard, S., Stutts, J., Courtney, T. K., & Connor, J. L. (2011). The link between fatigue and safety. *Accident Analysis & Prevention*, *43*(2), 498–515. https://doi.org/10.1016/j.aap.2009.11.011

Wilson, M. K., Ballard, T., Strickland, L., Amy Boeing, A., Cham, B., Griffin, M. A., & Jorritsma, K. (2021). Understanding fatigue in a naval submarine: Applying biomathematical models and workload measurement in an intensive longitudinal design. *Applied Ergonomics*, *94*, 103412. https://doi.org/10.1016/j.apergo.2021.103412

Wilson, M. K., Boag, R. J., & Strickland, L. (2019). All models are wrong, some are useful, but are they reproducible? Commentary on Lee et al. (2019). *Computational Brain & Behavior*, *2*(3–4), 239–241. https://doi.org/10.1007/s42113-019-00054-x

Wilson, M. K., Strickland, L., & Ballard, T. (2020). FIPS: An R Package for Biomathematical Modelling of Human Fatigue Related Impairment. *Journal of Open Source Software*, *5*(51), 2340. https://doi.org/10.21105/joss.02340

Young, M. S., & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(3), 365–375. https://doi.org/10.1518/0018720024497709