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

Micah K. Wilson[a*], Luke Strickland[a], Timothy Ballard[b], & Mark A. Griffin[a]

[a]*Future of Work Institute, Curtin University, Perth, Australia;* [b] *School of Psychological Science, University of Queensland, Brisbane, Australia*

Address correspondence to Micah K. Wilson, Curtin University, 78 Murray Street, Perth, Western Australia, 6000 (micah.k.wilson@curtin.edu.au).

Funding details: This research was supported by funding from the Australian Defence Science and Technology Group under MyIP:9079; Forrest Research Foundation Prospect Fellowship awarded to Micah K. Wilson.

Contribution note: Micah K. Wilson and Luke Strickland contributed to the work equally.

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

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 paper 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 that 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, practitioners, 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: cognitive modelling; alertness; performance; sleepiness; prediction

Relevance to human factors/ergonomics theory

Mental fatigue poses risks to the operational safety and effectiveness of sociotechnical systems and can negatively impact human performance and health. To mitigate such risks, human factors researchers and practitioners frequently employ predictive biomathematical models of fatigue as part of fatigue management strategies. To better understand the dynamics of mental fatigue and improve our ability to mitigate against it, new avenues for improving the relevance, accuracy, and validity of fatigue models must be identified and evaluated.

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 neurobehavioural outcomes of fatigue (e.g., alertness or response time) using time of day and sleep/wake history (see Civil Aviation Safety Authority, 2014). For example, airlines utilise crew management systems that coordinate workforce allocation across the globe using projected fatigue, and militaries utilise BMMs to implement watchkeeping schedules that optimise operational readiness. The proliferation of BMM tools 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. These implementations have several uses, including to evaluate the relative fatigue risks of alternative work rosters, facilitate design of future technical systems (Boeing et al. 2020), and support accident investigations (Price and 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, M. James, and Caldwell 2019; Horrey et al. 2011). 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). Next-generation fatigue prediction could be achieved using several approaches. In this paper, we focus on development of models that are extensions or adaptations to BMMs, but acknowledge complementary machine-learning approaches (Section 2.4) and emerging neurophysiological models and latent variable estimation methods (Section 3.1). A focus on adapting BMMs is important because they are widely and routinely employed in fatigue-risk management systems by industry and continue to garner significant research interest. The need for improved BMMs has been recognised for some time by both industry and researchers (Flight Safety Foundation 2005; Hursh et al. 2004; Klerman and Hilaire 2007; Reifman 2004) with limitations of existing BMMs reviewed extensively (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; individualising them to specific operators or individuals (Liu et al. 2017; Reifman, Rajaraman, and Gribok 2007; Van Dongen, Bender, and Dinges 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 intensive environmental and workplace demands while maintaining high levels of performance and safety over lengthy missions (Bell et al. 2016; Cham et al. 2021).

The high enthusiasm for next-generation BMMs (e.g., Bendak and Rashid 2020; Civil Aviation Safety Authority 2014; Flynn-Evans et al. 2020; Stone et al. 2020) has been maintained by advances in fatigue science, machine learning, and sensor technologies which passively detect human fatigue. 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 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. Dinges (2004, A182) concluded that “Most current models of fatigue and its effects on performance appear to be more descriptive curvefitting, than theoretically driven, hypothesis-generating, data-organising, mathematical approaches”. There have been few changes in this regard since Dinges’s assessment in 2004.

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 human factors practitioners and researchers can have regarding their realistic effectiveness, in turn, fostering new approaches to safety optimisation and research.

We begin the paper by reviewing the key characteristics of next-generation models, focusing on tuning model parameters using field-derived data (Section 2.1); individual-level parameter tuning and real-time fatigue prediction (Section 2.2); and extensions to modelling algorithms to account for additional factors that influence fatigue (Section 2.3). We then introduce and review alternative emerging methods, including machine-learning, that may augment or provide alternatives to BMMs (Section 2.4). Throughout these sections, we summarise research progress, identify practical and theoretical constraints limiting real-world applications, and utilise simulations and quantitative examples to explicate our arguments. We conclude the paper with a general summary of our findings and discuss the 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. Current generation BMMs are largely developed in controlled laboratory settings using samples of convenience. There is consensus that BMMs lack extensive validation in many operational contexts and bear only a coarse relationship with real-world risk (Dawson 2012; Dawson, Darwent, and Roach 2017; Gander et al. 2011; James et al. 2018; Reifman, Rajaraman, and Gribok 2007; Riedy, Roach, and Dawson 2020). The second capability is an important related case of model tailoring known as individualisation — that is, the capability to *individualise* predictions to specific employees. 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, which focus on group-level average predictions, are 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 (Horrey et al. 2011). Current research has focused predominantly on the influence of pharmaceutical fatigue countermeasures (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, assess the limitations of BMMs for next-generation modelling, and provide readers with practical future research directions.

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, Darwent, and Roach 2017; Dean et al. 2007; Williamson et al. 2011). Indeed, many operational contexts involve challenges that make the assumptions of default BMM parameterisations inappropriate. For instance, in the submarine context (a setting we examine later in this paper), the lack of exposure to natural light sources and artificial sleep-wake patterns (demanded by rostering constraints) can disrupt circadian processes and rhythmicity, potentially altering the predictive contribution of the circadian processes in BMMs (Cham et al. 2021; Guo et al. 2020; Sandal, Leon, and Palinkas 2006). Similarly, sleep quality can be disrupted by environmental factors like motion or ambient noise, potentially influencing the homeostatic recovery rate (Beare et al. 1981; Guo et al. 2020). Thus, using BMMs based on laboratory data may limit the accuracy of workplace fatigue predictions. In turn, this may compromise risk mitigation efforts when performing forward-scheduling or mission planning (Flynn-Evans et al. 2020; Reifman, Rajaraman, and Gribok 2007; Wilson et al. 2021).

There are several ways that BMMs could be augmented or adjusted to better represent the dynamics of work populations of interest. 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 one way to *tailor* or *tune* a BMM is to adjust model parameters to better describe fatigue measurements observed from a specific work population. Tailoring BMMs requires representative fatigue data that models can be trained with, yet obtaining appropriate data that reliably improves BMM predictions is challenging. One solution is to conduct high-fidelity laboratory studies with an employee sample and simulate their expected workplace conditions and demands, for example by using synthetic task environments (Flynn-Evans et al. 2020; Gonzalez, Vanyukov, and Martin 2005). This approach allows experimental control over the exposure to natural light, timing and duration of sleep, and the nature of work-representative tasks. This method greatly improves external validity while retaining the experimental control required for model estimation and development (e.g., Vital-Lopez, Doty, and Reifman 2021).

Applying an experimental approach may not be feasible in many industries due to the constraints in faithfully representing the relevant factors affecting fatigue in a workplace with laboratory resources. For many industries, it is costly to retain experts in laboratories for the durations necessary to ascertain fatigue trajectories, which can be on the order of days (e.g., in maritime domains, van Leeuwen et al. 2020). Therefore, an appealing alternative is to capture individuals' fatigue and sleep data directly in the work environment or operational context. The benefits of this approach are: 1) it offers the best chance of capturing the strain and recovery dynamics directly as they unfold in response to the environmental stressors which influence the underpinning neurobiological fatigue processes; and 2) field fatigue measurement is essential for 'real time' prediction, in which fatigue forecasts are updated based on incoming field data.

The challenges to field estimation must be considered. To successfully estimate BMM parameters from field data in workplace settings, data collection must be minimally invasive, and yet comprehensive enough to identify the complex non-linear dynamics specified by BMMs. Given the many possible processes that underlie fatigue in the field, it is probable that fatigue and sleep measurements are affected by 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 assurances (e.g., Ramakrishnan et al. 2016). However, cross-validation does not ensure BMMs *measure* the underlying fatigue processes of an individual or group, a goal which has been pursued in the literature (Ramakrishnan et al. 2015). Cross-validation speaks only to predictive accuracy for a particular set of data, and therefore provides no guarantees that a set of BMM parameter estimates will generalise to data ranges (e.g., sleep schedules) outside of those that have been tested. Indeed, St Hilaire et al. (2017) found substantial mis-fit when applying existing pre-trained BMMs to predict fatigue in a study involving chronic variable sleep deficiency, suggesting that the BMM parameter estimates may have been overfit to the sleep schedules from the studies in which they were initially 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 informs 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. 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 generation, 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 about the most likely parameter estimates and 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 distance from the centre line. To characterise the posterior mean parameter estimates, we also present mean absolute bias error (Equation 1) and normalised root mean square error (Equation 2):

$$\mathrm{MABE} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left|\widehat{\theta_i} - \theta_i\right| \tag{1}$$

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta_i - \widehat{\theta_i}\right)^2}}{\max\theta - min(\theta)} \cdot 100 \tag{2}$$

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.

[Insert Figure 1 approximately here]

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 assumes there are no additional factors (e.g., workload or fatigue countermeasures) that bear systematic influence on fatigue, which would be violated in naturalistic environments. We 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 sufficiently, 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). Parameters requiring most attention were the time constants of the homeostatic process. These recovered poorly, implying that the time course of the fatigue response to sleep/wake was difficult to estimate under even ideal conditions. As identifying the time course of the homeostatic process is one of the primary interests of field estimation, this suggests there may be limited 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 individualisation and extension.

2.2. Model Individualisation and Real-Time Prediction

The so called "Holy Grail in fatigue and performance modeling" is fatigue prediction individualisation through tailoring model parameterisations to each person (Reifman 2004, A177). The theory of individualisation 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 parameterisations for each person uniquely. Consistent with group-level estimation, parameters can be estimated using observations 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, enabling reactive identification of workplace fatigue risk (e.g., see Liu et al., 2017). Although this approach requires considerable amounts of data per person (both sleep and fatigue observations), there are strong arguments that characterising fatigue dynamics at the individual-level may improve model performance.

From a practical perspective, in many safety-critical workplaces it is most useful to obtain fatigue projections for specific employees over a period of hours to a few days (fitness to work), rather than knowing if an average employee would stay below fatigue safety thresholds given a particular roster. In complex 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). Thus, real-time fatigue forecasts made possible by individualisation promise improved tactical decision making (e.g., ideal time to execute mission scenarios) and crew rotation decisions (e.g., which staff are 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, and vulnerability to sleep deprivation (Chua et al. 2019). Early research has indicated individualised models are possible (Dawson et al. 2011; Ramakrishnan et al. 2015) with uncertainty regarding which parameters should be considered as stable trait differences, relative to state differences that fluctuate within-person (Ramakrishnan et al. 2015; Van Dongen et al. 2007). A key benefit of individualisation is that BMM parameters can be informed by measures other than the performance criterion. For instance, recent circadian modelling research has indicated that lighting conditions bear strong predictive influence over circadian entrainment and sleep timing (Papatsimpa et al. 2021; Phillips et al. 2019). Thus, BMM circadian phase parameters could be informed from actigraphy and photometry data passively (see Brown et al. 2021) to reduce model complexity and improve predictive accuracy. This means parameter estimation using behavioural data would only need to be conducted for a subset of model parameters.

Despite the appeal of BMM individualisation, there are several notable implementation challenges. The parameter recovery issues outlined in Section 2.1. apply even more strongly when the requirement is to tune BMM parameters to the sparser individual-level observations. Consequently, existing individualised 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 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 is uniquely associated with the processes assumed within the BMM (unlike laboratory contexts where many factors are controlled) (Reifman, Rajaraman, and Gribok 2007). For example, if employees face high levels of work-induced fatigue, and this is not 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 selected 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 criterion variable used in the model.

In summary, individualising BMMs is a priority for next-generation modelling and promises many potential benefits. Although fitting BMMs to the behavioural data of individuals holds some promise towards this goal, it is constrained by substantial data limitations. Future approaches to individualisation, particularly involving field data, are likely to greatly benefit from incorporating other individualised 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 (Desmond and 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 models including processes for chronic sleep restriction (e.g., Rajdev et al. 2013). Research has focused predominantly on the influence of pharmaceutical fatigue countermeasures (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, the benefits and barriers involved, and required assumptions/decisions, we detail a complete example of the model extension process using the salient example of how workload may modulate fatigue. It is uncontroversial that fatigue is influenced by work factors, such as shift duration and workload (Desmond and Hancock 2001; Grech et al. 2009; Wilson et al. 2021). Consequently, there has been much discussion of extending BMMs to model 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 work-related fatigue accrues over time spent working, at a rate proportional to the task-imposed workload and current homeostatic fatigue. Honn et al. (2016) incorporated a similar approach into the McCauley state-space model.

Equation 3 specifies how workload-related fatigue accrues during work time, with the speed of accrual at each moment dependent on homeostatic pressure. Equation 4 specifies how recovery from work-related fatigue follows an exponential function. The computational implementation of the model is available (see data availability statement).

$$D_t = D_0 + \gamma \times \int_0^t \max(S_t, 0)\, dt \tag{3}$$

$$D_t = D_l - e^{-\frac{t}{\tau_r}}(D_l - D_0) \tag{4}$$

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. The parameters controlling the workload process, such as γ, could be estimated by fitting model predictions to fatigue observations obtained during work conditions (see also Honn et al., 2016). It would also be possible to model the fatigue accrual associated with specific levels of task demand by scaling the γ parameter as a function of task-load estimates or subjective workload ratings. Figure 2 shows the impact of adding this workload-related fatigue module onto the basal fatigue unified model predictions for a 0900-1700 work roster. The plot uses 16 different γ values, which is equivalent to plotting the process under a range of different workload demands.

[Insert Figure 2 approximately here]

The workload model described is representative of the typical BMM extension process. 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, research accumulates results towards a well validated model component with theoretical rigor (e.g., the sleep-inertia component of the three-process model, Åkerstedt and Folkard 1997). Validation requires many of the same considerations as those for parameter estimation discussed in section 2.1. For example, to determine appropriate parameterisations of our model, researchers would need to conduct a controlled laboratory study that systematically manipulated workload. It must also be determined whether work-related fatigue elicited using laboratory tasks generalises to 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 neglect possible parameter trade-offs, raising concerns of model identifiability. Indeed, the recovery behaviour of even baseline BMMs (Section 2.1.1.) suggests significant challenges exist in freely estimating extended BMMs.

A theory-based challenge associated with parametric model extension is managing the increasing model complexity during exploration and validation. Unlike more descriptive conventional statistical approaches, BMMs precisely specify the functional form of their component processes, and the relationships between these processes. Extant knowledge and theory 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 alternatives. For instance, the workload model we introduced above may need to account for situations of underload induced fatigue (Shultz, Wang, and Olson 2010; Young and Stanton 2002) or for the known "carry over" effects of high workload situations on subsequent sleep (Crain, Brossoit, and Fisher 2018). Each such point of additional complexity should be weighed against improvement prediction accuracy. Further, 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 demands 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 support the development of safety-promotion strategies (Reifman et al. 2016; 2019). Similarly, imperfect models of workload provide practitioners a method to identify possible high-risk roster scenarios and formalise assumptions in a manner that would otherwise remain as qualitative verbal theory (Ballard et al. 2021). Thus, these techniques offer practical benefits to practitioners, but are unlikely to provide immediate step-changes in predictive accuracy, theoretical advancement, or real-time operational safety.

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 alternative methodological approaches, and how they may help address these goals. A frequently noted direction is the adoption of machine learning, a class of data-driven statistical techniques that allow computers to learn from data and generate predictions without the need for explicit model structures or instructions (Jordan and Mitchell 2015).

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 make predictions without requiring an a priori model structure (see also Breiman 2001). A benefit of machine-learning approaches is that, in principle, they can incorporate any number of predictors (Jordan and 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 "hybrid methods" in which BMMs are embedded within neural network architectures to improve prediction. Machine-learning can also enable integration of real-time physiological indicators of fatigue, such as cardiovascular state (Aryal, Ghahramani, and Becerik-Gerber 2017; Hu and Lodewijks 2020), although these approaches have been argued to have limited utility and validity (Dawson, Searle, and Paterson 2014). Given sufficiently mature computational infrastructure, machine-learning approaches could predict fatigue during operations, potentially recommending interventions when fatigue risk is high. The prospect of deploying such systems in workplace environments is increasingly possible due to advances in statistical methods, data storage, and computational power.

It is timely to begin identifying and validating applications of machine-learning methods into the fatigue prediction toolkit. Even for researchers who continue to pursue BMM extension, machine-learning methods can indirectly integrate with BMMs. For example, machine-learning can help generate estimates of individuals' sleep quality and quantity from wearable technologies (Lewicke et al. 2008; Piotrowski and Szypulska 2017). Sundararajan et al. (2021) recently applied 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 also 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 (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 do have substantial data requirements, and provide only 'black box' predictions that cannot be easily decomposed into the underlying processes (e.g., circadian rhythm). 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 applying to different workplaces). Cochrane et al. (2021) used an ensemble machine-learning model to predict the effects of sleep-loss in a forced desynchrony protocol dataset. However, the data requirements for accurate prediction were significant, requiring 10-minute PVT administration every 2-8 hours, and it remained 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. This is likely the most promising direction for next-generation 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 the naval submarine context, which revealed that even under optimistic modelling conditions, estimation using field data was likely to produce highly uncertain estimates for at least several critical 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 individualisation. While we noted some promising advancements have been made in laboratory contexts, overall, individualisation is constrained to tightly controlled estimation of a subset of parameters. Thirdly, we examined the practice of parametric BMM extension, focusing on extending a BMM to incorporate workload. Our example highlighted the substantial challenges associated with extending BMMs, including the requirement for precise theoretically informed mappings between any new BMM process and the underlying fatigue function (e.g., sleepiness) and interactions with other model processes. We concluded next-generation

BMMs involving model extensions are likely prohibitively difficult to accurately specify, and extremely challenging to validate.

It is crucial to emphasise 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 or individuals, or models that can incorporate domain-relevant variables such as workload. However, our review highlights significant barriers to achieving these goals with BMMs alone. In the following section, we provide recommendations to guide researchers and model developers toward what we believe are fruitful avenues for advancing the theory and practice of fatigue modelling, particularly as it applies to human factors.

3.1. Recommendations for Advancing Fatigue Modelling

3.1.1. Expanding Theoretical Frameworks of Fatigue and Performance

The equations underlying the BMM approach only capture a proportion of the complex dynamics underlying fatigue. Moving forward, fatigue science must emphasise multi-model approaches and shift how fatigue is measured and modelled to match advances in other computational modelling fields. Several challenges are clear. There is a need to better specify the relationship between the predicted model output and the actual performance criterion it intends to capture. Our recovery analysis revealed that the behavioural response to fatigue is the element that BMMs are least effective in indexing (i.e., homeostatic process). It follows that this is an important area in which alternative frameworks and approaches could help address.

Historically, fatigue researchers have relied on coarse data (e.g., response time, lapse rates, and subjective scales) for model validation and theoretical innovation. This reliance on coarse measures has restricted practitioners and researchers to only envision predictions in terms of those same coarse measures. For example, validation efforts for commercial models have generally only linked coarse predictions of measures (based on sleep opportunity) with historical safety incident rates (Hursh et al. 2006). Yet these outputs bear weak relevance to complex workplace task performance (Williamson et al. 2011). Solutions have been proposed, including that models be calibrated against task performance metrics obtained from real-world scenarios, or representative simulations (Reifman, Rajaraman, and Gribok 2007). Such efforts would improve the ecological validity of model outputs but do little for generalisation. A promising pathway to ensure performance predictions can generalise across scenarios and contexts may be to model the latent mechanisms underlying performance.

Although BMMs are physiologically inspired (e.g., Borberly, 1982), they do not explicitly model the physiological processes underlying fatigue. To address this, newer neurobiological models have been proposed that more directly relate behavioural model predictions (e.g., PVT performance) to the neurobiological mechanisms underpinning the sleep-wake drive (Fulcher, Phillips, and Robinson 2010; Phillips, Klerman, and Butler 2017; St. Hilaire et al. 2016). There are theoretical and practical advantages to incorporating prediction from these approaches. For example, St. Hilaire et al. (2016) developed the “adenosine model” which simulates cerebral extracellular adenosine dynamics (and receptor concentrations) to predict PVT performance under conditions of chronic variable sleep deficiency (insufficient and inconsistent night-to-night sleep duration). They compared predictive accuracy of the adenosine model to four BMMs (including the unified model) for modelling performance impairment under these complex sleep cycles. Although no model provided a single best account of all data features, only the adenosine model successfully predicted the magnitudes of fatigue-related performance impairment and recovery under the variable sleep deficiency. Concerningly, the results suggested conventional BMMs *overestimated* the recovery achieved from 8-10 hours sleep following chronic sleep deficiency, which may pose operational safety concerns if not considered. This highlights the importance of adopting multi-model approaches in fatigue-risk assessment, particularly those that are based on neurologically plausible mechanisms.

3.1.1.1. Computational Cognitive Models. Understanding the latent constructs underpinning the behavioural performance changes in response to fatigue requires dynamic models of behaviour. There have been major developments in computational cognitive models that specify in detail the processes underlying task performance. These models provide a means to quantify the effects of fatigue on performance and account for the processes influencing criterion variables. For instance, PVT metrics such as number of lapses and mean RT have ambiguous mappings to underlying cognitive processes (Chua et al. 2019; Veksler and Gunzelmann 2018). Consequently, it is unclear whether fatigue increases RT and lapses because individuals process information less efficiently when fatigued, or because they are responding more cautiously (i.e., require more evidence to respond). These competing explanations have direct implications for the safety profile of tasks when fatigued, but can be adjudicated using evidence accumulation models which use response choice and response time data to measure the latent cognitive constructs such as processing speed and caution. Evidence accumulation models have been applied to PVT performance (Chavali, Riedy, and Van Dongen 2017; Ratcliff and Van Dongen 2011), and integrated with BMMs to a limited extent (Walsh, Gunzelmann, and Van Dongen 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 broader ACT-R cognitive architecture (Gunzelmann, M. James, and Caldwell 2019) offering opportunities to generalise performance predictions.

Sophisticated computational models of behaviour help to differentiate causes of the fatigue response. BMMs, including the workload extensions, generally ground fatigue accumulation as largely a homeostatic driven process. This research is founded on findings with the PVT, which was initially selected as it is sensitive sleep-loss. However, 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 impact on performance. Cognitive models can enable researchers to quantitatively test the influences of such factors.

3.1.1.2. Dynamic Longitudinal Models. A crucial development for understanding the latent dynamics underlying fatigue, particularly when considering field measurement, is dynamic models. Dynamic structural equation modelling (Asparouhov, Hamaker, and Muthén 2018; Driver and Voelkle 2018) allows researchers to model how latent variables evolve and relate to each other over time (i.e., auto- and cross-regressive effects). Thus, they can inform how fatigue interacts and coevolves with individual, environmental, and work-related factors over time. These approaches can also address questions of temporal causality between measures, and unlike BMMs do not require exact mathematical specifications of interactions between factors. Dynamic models may potentially detect longer-term ‘knock on’ effects of workload to sleep quality and quantity, and long-term burnout (Crain, Brossoit, and Fisher 2018; Wilson et al. 2021).

Many of the variables related to fatigue are state dependent, meaning the causal relationships among variables change with different states of the system (Chang, Ushio, and Hsieh 2017). For instance, a sustained level of high workload may cause fatigue, but high levels of fatigue may reduce cognitive capacity causing higher workload (Wilson et al. 2021). Methods such as *empirical dynamic modelling* (Chang et al. 2017) may offer a means to decompose such complex interdependencies in causal systems, and improve our ability to mitigate risk in safety-critical workplaces.

3.1.2. Improve Open Science Practices

Open science practices are increasingly integral to achieving robust science (Munafò et al. 2017). The fatigue sciences require significant efforts be placed towards computational reproducibility and transparency. Presently, most fatigue prediction solutions are closed source and proprietary, and in many cases, independent replication or extension of the work reported in scientific articles is impractical or impossible. In cases where BMM formulae are provided, the respective computational implementations generally 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 (see R. D. Peng 2011; Wilson, Boag, and Strickland 2019). Initial steps towards computational reproducibility have been taken with the development of *2B-Alert Web*, an open-access application that provides graphical BMM predictions (Reifman et al. 2016), and the release of the open-source R package for BMMs, *FIPS*, which provides sleep and fatigue data structures and BMM implementations (Wilson, Strickland, and Ballard 2020). Such efforts encourage cumulative science and mitigate research fragmentation.

There is also consensus that open data practices can accelerate methodological innovation and scientific discovery (Gewin 2016; Meyer 2018). For fatigue prediction, open data can enable new models to be evaluated on benchmark datasets and thereby avoid the situation where BMMs are extended using new datasets, without establishing if it still adequately fits the data from which the original models were developed (Reifman and Gander 2004). Improved validity is inherently valuable for the employees and organisations that fatigue prediction models are intended to benefit. Indeed, in reviewing BMMs, the Flight Safety Foundation (2005) emphasised their desire for improved data sharing practices in fatigue science and stressed the importance of transparent science for building industry trust. Admittedly, individuals face barriers to open science practices, including data privacy and intellectual property concerns. Practical considerations such as funding have long challenged the advancement of fatigue prediction methods (Akerstedt, Folkard, and Portin 2004). However, this does not undermine the value of the aforementioned practices. Engagement in open science is essential to overcome the barriers to achieving next-generation model features.

3.2 Conclusions

BMMs have been instrumental in advancing evidence-based fatigue management strategies in safety-critical contexts. Significant interest remains in the development of next-generation BMMs capable of providing tailored and more accurate fatigue predictions. This paper reviewed and analysed several key directions underpinning the development of next-generation models and has revealed there are significant challenges to realise the 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.

Data availability statement

Full information pertaining to analyses conducted in the manuscript, including the code used to conduct modelling, are stored in the manuscript's open science framework repository and are accessible via https://osf.io/yurvx/

Indication of Figures

Two figures associated with paper.

Declarations of interest

None

Word Count

10341

References

Åkerstedt, Torbjörn, and Simon Folkard. 1997. 'The Three-Process Model of Alertness and Its Extension to Performance, Sleep Latency, and Sleep Length'. *Chronobiology International* 14 (2): 115–23. https://doi.org/10.3109/07420529709001149.

Akerstedt, Torbjörn, Simon Folkard, and Christian Portin. 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.

Aryal, Ashrant, Ali Ghahramani, and Burcin Becerik-Gerber. 2017. 'Monitoring Fatigue in Construction Workers Using Physiological Measurements'. *Automation in Construction* 82 (October): 154–65. https://doi.org/10.1016/j.autcon.2017.03.003.

Asparouhov, Tihomir, Ellen L. Hamaker, and Bengt Muthén. 2018. 'Dynamic Structural Equation Models'. *Structural Equation Modeling: A Multidisciplinary Journal* 25 (3): 359–88. https://doi.org/10.1080/10705511.2017.1406803.

Ballard, Timothy, Hector Palada, Mark Griffin, and Andrew Neal. 2021. 'An Integrated Approach to Testing Dynamic, Multilevel Theory: Using Computational Models to Connect Theory, Model, and Data'. *Organizational Research Methods* 24 (2): 251–84. https://doi.org/10.1177/1094428119881209.

Beare, Arthur N., Kenneth R. Bondi, Robert J. Biersner, and Paul Naitoh. 1981. 'Work and Rest on Nuclear Submarines'. *Ergonomics* 24 (8): 593–610. https://doi.org/10.1080/00140138108924882.

Bell, Suzanne T., David M. Fisher, Shanique G. Brown, and Kristin E. Mann. 2016. 'An Approach for Conducting Actionable Research With Extreme Teams': *Journal of Management*, June. https://doi.org/10.1177/0149206316653805.

Bendak, Salaheddine, and Hamad S. J. Rashid. 2020. 'Fatigue in Aviation: A Systematic Review of the Literature'. *International Journal of Industrial Ergonomics* 76 (March): 102928. https://doi.org/10.1016/j.ergon.2020.102928.

Boeing, Alexandra Amy, Karina Jorristma, Mark A Griffin, and Sharon K Parker. 2020. 'Surfacing the Social Factors Early: A Sociotechnical Approach to the Design of a Future Submarine'. *Australian Journal of Management* 45 (3): 527–45. https://doi.org/10.1177/0312896220920338.

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

Brown, Lindsey S., Melissa A. St Hilaire, Andrew W. McHill, Andrew J. K. Phillips, Laura K. Barger, Akane Sano, Charles A. Czeisler, Francis J. Doyle, and Elizabeth B. Klerman. 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, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. 'Stan: A Probabilistic Programming Language'. *Journal of Statistical Software, Articles* 76 (1): 1–32. https://doi.org/10.18637/jss.v076.i01.

Cham, Belinda S., Alexandra A. Boeing, Micah K. Wilson, Mark A. Griffin, and Karina Jorritsma. 2021. 'Endurance in Extreme Work Environments'. *Organizational Psychology Review*, April, 20413866211006440. https://doi.org/10.1177/20413866211006441.

Chang, Chun-Wei, Masayuki Ushio, and Chih-hao Hsieh. 2017. 'Empirical Dynamic

Modeling for Beginners’. *Ecological Research* 32 (6): 785–96. https://doi.org/10.1007/s11284-017-1469-9.
Chavali, Venkata P., Samantha M. Riedy, and Hans P.A. Van Dongen. 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, Eric Chern-Pin, Jason P. Sullivan, Jeanne F. Duffy, Elizabeth B. Klerman, Steven W. Lockley, Bruce S. Kristal, Charles A. Czeisler, and Joshua J. Gooley. 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, Courtney, Demba Ba, Elizabeth B. Klerman, and Melissa A. St. Hilaire. 2021. ‘An Ensemble Mixed Effects Model of Sleep Loss and Performance’. *Journal of Theoretical Biology* 509 (January): 110497. https://doi.org/10.1016/j.jtbi.2020.110497.
Crain, Tori L., Rebecca M. Brossoit, and Gwenith G. Fisher. 2018. ‘Work, Nonwork, and Sleep (WNS): A Review and Conceptual Framework’. *Journal of Business and Psychology* 33 (6): 675–97. https://doi.org/10.1007/s10869-017-9521-x.
Dawson, Drew. 2012. ‘Fatigue Research in 2011: From the Bench to Practice’. *Accident Analysis & Prevention*, Managing Fatigue in Transportation, Resources and Health8th International ConferenceFremantle, Western Australia21-24 March 2011, 45 (March): 1–5. https://doi.org/10.1016/j.aap.2011.09.016.
Dawson, Drew, David Darwent, and Gregory D. Roach. 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 (February): 469–73. https://doi.org/10.1016/j.aap.2015.11.032.
Dawson, Drew, Y. Ian Noy, Mikko Härmä, Torbjorn Åkerstedt, and Gregory Belenky. 2011. ‘Modelling Fatigue and the Use of Fatigue Models in Work Settings’. *Accident Analysis & Prevention* 43 (2): 549–64. https://doi.org/10.1016/j.aap.2009.12.030.
Dawson, Drew, Amelia K. Searle, and Jessica L. Paterson. 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): 141–52. https://doi.org/10.1016/j.smrv.2013.03.003.
Dean, Dennis A., Adam Fletcher, Steven R. Hursh, and Elizabeth B. Klerman. 2007. ‘Developing Mathematical Models of Neurobehavioral Performance for the “Real World”’. *Journal of Biological Rhythms* 22 (3): 246–58. https://doi.org/10.1177/0748730407301376.
Desmond, Paula A., and Peter A. Hancock. 2001. ‘Active and Passive Fatigue States.’ In *Stress, Workload, and Fatigue.*, 455–65. Human Factors in Transportation. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
Dinges, D.F. 2004. ‘Critical Research Issues in Development of Biomathematical Models of Fatigue and Performance’. *Aviation Space and Environmental Medicine* 75 (3): A181–91.
Driver, Charles C., and Manuel C. Voelkle. 2018. ‘Hierarchical Bayesian Continuous Time Dynamic Modeling.’ *Psychological Methods* 23 (4): 774–99. 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, Erin E., Crystal Kirkley, Millennia Young, Nicholas Bathurst, Kevin Gregory, Verena Vogelpohl, Albert End, Steven Hillenius, Yvonne Pecena, and Jessica J. Marquez. 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.
Fulcher, B. D., A. J. K. Phillips, and P. A. Robinson. 2010. ‘Quantitative Physiologically Based Modeling of Subjective Fatigue during Sleep Deprivation’. *Journal of Theoretical Biology* 264 (2): 407–19. https://doi.org/10.1016/j.jtbi.2010.02.028.
Gander, Philippa, Laurence Hartley, David Powell, Philippe Cabon, Edward Hitchcock, Ann Mills, and Stephen Popkin. 2011. ‘Fatigue Risk Management: Organizational Factors at the Regulatory and Industry/Company Level’. *Accident Analysis & Prevention*, Advancing Fatigue and Safety Research, 43 (2): 573–90. https://doi.org/10.1016/j.aap.2009.11.007.
Gewin, Virginia. 2016. ‘Data Sharing: An Open Mind on Open Data’. *Nature* 529 (7584): 117–19. https://doi.org/10.1038/nj7584-117a.
Gonzalez, Cleotilde, Polina Vanyukov, and Michael K. Martin. 2005. ‘The Use of Microworlds to Study Dynamic Decision Making’. *Computers in Human Behavior* 21 (2): 273–86. https://doi.org/10.1016/j.chb.2004.02.014.
Grech, Michelle R., Andrew Neal, Gillian Yeo, Michael Humphreys, and Simon Smith. 2009. ‘An Examination of the Relationship between Workload and Fatigue within and across Consecutive Days of Work: Is the Relationship Static or Dynamic?’ *Journal of Occupational Health Psychology* 14 (3): 231–42. https://doi.org/10.1037/a0014952.
Gunzelmann, Glenn, Stephen M. James, and Jo Lynn Caldwell. 2019. ‘Basic and Applied Science Interactions in Fatigue Understanding and Risk Mitigation’. In *Progress in Brain Research*, 246:177–204. Elsevier. https://doi.org/10.1016/bs.pbr.2019.03.022.
Guo, Jin-Hu, Xiao-Hong Ma, Huan Ma, Yin Zhang, Zhi-Qiang Tian, Xin Wang, and Yong-Cong Shao. 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.
Honn, Kimberly A., Brieann C. Satterfield, Peter McCauley, J. Lynn Caldwell, and Hans P. A. Van Dongen. 2016. ‘Fatiguing Effect of Multiple Take-Offs and Landings in Regional Airline Operations’. *Accident Analysis & Prevention* 86 (January): 199–208. https://doi.org/10.1016/j.aap.2015.10.005.
Horrey, William J., Y. Ian Noy, Simon Folkard, Stephen M. Popkin, Heidi D. Howarth, and Theodore K. Courtney. 2011. ‘Research Needs and Opportunities for Reducing the Adverse Safety Consequences of Fatigue’. *Accident Analysis & Prevention*, Advancing Fatigue and Safety Research, 43 (2): 591–94. https://doi.org/10.1016/j.aap.2010.01.014.
Hu, Xinyun, and Gabriel Lodewijks. 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*, January, S0022437519306735. https://doi.org/10.1016/j.jsr.2019.12.015.
Hursh, Steven R., Thomas G. Raslear, A. Scott Kaye, and Joseph F. Fanzone. 2006. ‘Validation and Calibration of a Fatigue Assessment Tool for Railroad Work

Schedules, Summary Report: (736302011-001)’. American Psychological Association. https://doi.org/10.1037/e736302011-001.
Hursh, Steven R., Daniel P. Redmond, Michael L. Johnson, David R. Thorne, Gregory Belenky, Thomas J. Balkin, William F. Storm, James C. Miller, and Douglas R. Eddy. 2004. ‘Fatigue Models for Applied Research in Warfighting’. *Aviation, Space, and Environmental Medicine* 75 (3): A44–53.
James, Francine O., Lauren B. Waggoner, Patricia M. Weiss, P. Daniel Patterson, J. Stephen Higgins, Eddy S. Lang, and Hans P. A. Van Dongen. 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., and T. M. Mitchell. 2015. ‘Machine Learning: Trends, Perspectives, and Prospects’. *Science* 349 (6245): 255–60. https://doi.org/10.1126/science.aaa8415.
Klerman, Elizabeth B., and Melissa St. Hilaire. 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.
Leeuwen, Wessel M. A. van, Claire Pekcan, Mike Barnett, and Göran Kecklund. 2020. ‘Mathematical Modelling of Sleep and Sleepiness under Various Watch Keeping Schedules in the Maritime Industry’. *Marine Policy*, October, 104277. https://doi.org/10.1016/j.marpol.2020.104277.
Lewicke, A., E. Sazonov, M.J. Corwin, M. Neuman, and S. Schuckers. 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): 108–18. https://doi.org/10.1109/TBME.2007.900558.
Liu, Jianbo, Sridhar Ramakrishnan, Srinivas Laxminarayan, Thomas J. Balkin, and Jaques Reifman. 2017. ‘Real-Time Individualization of the Unified Model of Performance’. *Journal of Sleep Research* 26 (6): 820–31. https://doi.org/10.1111/jsr.12535.
Meyer, Michelle N. 2018. ‘Practical Tips for Ethical Data Sharing’. *Advances in Methods and Practices in Psychological Science* 1 (1): 131–44. https://doi.org/10.1177/2515245917747656.
Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. ‘A Manifesto for Reproducible Science’. *Nature Human Behaviour* 1 (1): 1–9. https://doi.org/10.1038/s41562-016-0021.
Noy, Y. Ian, William J. Horrey, Stephen M. Popkin, Simon Folkard, Heidi D. Howarth, and Theodore K. Courtney. 2011. ‘Future Directions in Fatigue and Safety Research’. *Accident Analysis & Prevention*, Advancing Fatigue and Safety Research, 43 (2): 495–97. https://doi.org/10.1016/j.aap.2009.12.017.
Papatsimpa, C., L. J. M. Schlangen, K. C. H. J. Smolders, J.-P. M. G. Linnartz, and Y. A. W. de Kort. 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, Henry T., Fethi Bouak, Wenbi Wang, Renee Chow, and Oshin Vartanian. 2018. ‘An Improved Model to Predict Performance under Mental Fatigue’.

Ergonomics, January, 1–16. https://doi.org/10.1080/00140139.2017.1417641.
Peng, R. D. 2011. ‘Reproducible Research in Computational Science’. *Science* 334 (6060): 1226–27. https://doi.org/10.1126/science.1213847.
Phillips, Andrew J. K., Elizabeth B. Klerman, and James P. Butler. 2017. ‘Modeling the Adenosine System as a Modulator of Cognitive Performance and Sleep Patterns during Sleep Restriction and Recovery’. Edited by Hans-Peter Landolt. *PLOS Computational Biology* 13 (10): e1005759. https://doi.org/10.1371/journal.pcbi.1005759.
Phillips, Andrew J. K., Parisa Vidafar, Angus C. Burns, Elise M. McGlashan, Clare Anderson, Shantha M. W. Rajaratnam, Steven W. Lockley, and Sean W. Cain. 2019. ‘High Sensitivity and Interindividual Variability in the Response of the Human Circadian System to Evening Light’. *Proceedings of the National Academy of Sciences*, May, 201901824. https://doi.org/10.1073/pnas.1901824116.
Piotrowski, Zbigniew, and Małgorzata Szypulska. 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, Jana M., and Bruce G. Coury. 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*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Rajaraman, Srinivasan, Andrei V. Gribok, Nancy J. Wesensten, Thomas J. Balkin, and Jaques Reifman. 2008. ‘Individualized Performance Prediction of Sleep-Deprived Individuals with the Two-Process Model’. *Journal of Applied Physiology* 104 (2): 459–68. https://doi.org/10.1152/japplphysiol.00877.2007.
———. 2009. ‘An Improved Methodology for Individualized Performance Prediction of Sleep-Deprived Individuals with the Two-Process Model’. *Sleep* 32 (10): 1377–92. https://doi.org/10.1093/sleep/32.10.1377.
Rajdev, Pooja, David Thorsley, Srinivasan Rajaraman, Tracy L. Rupp, Nancy J. Wesensten, Thomas J. Balkin, and Jaques Reifman. 2013. ‘A Unified Mathematical Model to Quantify Performance Impairment for Both Chronic Sleep Restriction and Total Sleep Deprivation’. *Journal of Theoretical Biology* 331 (August): 66–77. https://doi.org/10.1016/j.jtbi.2013.04.013.
Ramakrishnan, Sridhar, Wei Lu, Srinivas Laxminarayan, Nancy J. Wesensten, Tracy L. Rupp, Thomas J. Balkin, and Jaques Reifman. 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): 262–69. https://doi.org/10.1111/jsr.12272.
Ramakrishnan, Sridhar, Srinivasan Rajaraman, Srinivas Laxminarayan, Nancy J. Wesensten, Gary H. Kamimori, Thomas J. Balkin, and Jaques Reifman. 2013. ‘A Biomathematical Model of the Restoring Effects of Caffeine on Cognitive Performance during Sleep Deprivation’. *Journal of Theoretical Biology* 319 (February): 23–33. https://doi.org/10.1016/j.jtbi.2012.11.015.
Ramakrishnan, Sridhar, Nancy J. Wesensten, Thomas J. Balkin, and Jaques Reifman. 2016. ‘A Unified Model of Performance: Validation of Its Predictions across Different Sleep/Wake Schedules’. *Sleep* 39 (1): 249–62. https://doi.org/10.5665/sleep.5358.
Ratcliff, R., and H. P. A. Van Dongen. 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): 11285–90. https://doi.org/10.1073/pnas.1100483108.
Reifman, Jaques. 2004. 'Alternative Methods for Modeling Fatigue and Performance'. *Aviation, Space, and Environmental Medicine* 75 (3 Suppl): A173-180.
Reifman, Jaques, and Philippa Gander. 2004. 'Commentary on the Three-Process Model of Alertness and Broader Modeling Issues'. *Aviation, Space, and Environmental Medicine* 75 (3): 5.
Reifman, Jaques, Kamal Kumar, Nancy J. Wesensten, Nikolaos A. Tountas, Thomas J. Balkin, and Sridhar Ramakrishnan. 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): 2157–59. https://doi.org/10.5665/sleep.6318.
Reifman, Jaques, Srinivasan Rajaraman, and Andrei V. Gribok. 2007. 'Moving Towards Individualized Performance Models'. *Sleep* 30 (9): 1081–82. https://doi.org/10.1093/sleep/30.9.1081.
Reifman, Jaques, Sridhar Ramakrishnan, Jianbo Liu, Adam Kapela, Tracy J. Doty, Thomas J. Balkin, Kamal Kumar, and Maxim Y. Khitrov. 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, Samantha M, Gregory D Roach, and Drew Dawson. 2020. 'Sleep–Wake Behaviors Exhibited by Shift Workers in Normal Operations and Predicted by a Biomathematical Model of Fatigue'. *Sleep*, March, zsaa049. https://doi.org/10.1093/sleep/zsaa049.
Sandal, G. M., G. R. Leon, and L. Palinkas. 2006. 'Human Challenges in Polar and Space Environments'. *Reviews in Environmental Science and Bio/Technology* 5 (2): 281–96. https://doi.org/10.1007/s11157-006-9000-8.
Sense, Florian, Ryan Wood, Michael G. Collins, Joshua Fiechter, Aihua Wood, Michael Krusmark, Tiffany Jastrzembski, and Christopher W. Myers. 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, Kenneth S., Mo Wang, and Deborah A. Olson. 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.
St. Hilaire, Melissa A., Melanie Rüger, Federico Fratelli, Joseph T. Hull, Andrew J. K. Phillips, and Steven W. Lockley. 2016. 'Modeling Neurocognitive Decline and Recovery During Repeated Cycles of Extended Sleep and Chronic Sleep Deficiency'. *Sleep* 40 (1): zsw009. https://doi.org/10.1093/sleep/zsw009.
Stone, Julia E., Svetlana Postnova, Tracey L. Sletten, Shantha M. W. Rajaratnam, and Andrew J. K. Phillips. 2020. 'Computational Approaches for Individual Circadian Phase Prediction in Field Settings'. *Current Opinion in Systems Biology* 22 (August): 39–51. https://doi.org/10.1016/j.coisb.2020.07.011.
Sundararajan, Kalaivani, Sonja Georgievska, Bart H. W. te Lindert, Philip R. Gehrman, Jennifer Ramautar, Diego R. Mazzotti, Séverine Sabia, et al. 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, Hans P. A., Christopher G. Mott, Jen-Kuang Huang, Daniel J. Mollicone, Frederic D. McKenzie, and David F. Dinges. 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): 1129–43. https://doi.org/10.1093/sleep/30.9.1129.
Van Dongen, Hans P.A., Amy M. Bender, and David F. Dinges. 2012. 'Systematic Individual Differences in Sleep Homeostatic and Circadian Rhythm Contributions to Neurobehavioral Impairment during Sleep Deprivation'. *Accident Analysis & Prevention* 45 (March): 11–16. https://doi.org/10.1016/j.aap.2011.09.018.
Veksler, Bella Z., and Glenn Gunzelmann. 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, Francisco G, Tracy J Doty, and Jaques Reifman. 2021. 'Optimal Sleep and Work Schedules to Maximize Alertness'. *Sleep* 44 (11). https://doi.org/10.1093/sleep/zsab144.
Walsh, Matthew M., Glenn Gunzelmann, and Hans P. A. Van Dongen. 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, Ann, David A. Lombardi, Simon Folkard, Jane Stutts, Theodore K. Courtney, and Jennie L. Connor. 2011. 'The Link between Fatigue and Safety'. *Accident Analysis & Prevention*, Advancing Fatigue and Safety Research, 43 (2): 498–515. https://doi.org/10.1016/j.aap.2009.11.011.
Wilson, Micah K., Timothy Ballard, Luke Strickland, Alexandra Amy Boeing, Belinda Cham, Mark A. Griffin, and Karina Jorritsma. 2021. 'Understanding Fatigue in a Naval Submarine: Applying Biomathematical Models and Workload Measurement in an Intensive Longitudinal Design'. *Applied Ergonomics* 94 (July): 103412. https://doi.org/10.1016/j.apergo.2021.103412.
Wilson, Micah K., Russell James Boag, and Luke Strickland. 2019. 'All Models Are Wrong, Some Are Useful, but Are They Reproducible? Commentary on Lee et al. (2019)'. *Computational Brain & Behavior* 2 (3–4): 239–41. https://doi.org/10.1007/s42113-019-00054-x.
Wilson, Micah K., Luke Strickland, and Timothy Ballard. 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, Mark S., and Neville A. Stanton. 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): 365–75. https://doi.org/10.1518/0018720024497709.