Understanding Fatigue in a Naval Submarine: Applying Biomathematical Models and Workload Measurement in an Intensive Longitudinal Design

Micah Kate Wilson[1], Timothy Ballard[2], Luke Strickland[1], Alex Boeing[1], Belinda Cham[1], Mark Griffin[1], Karina Jorritsma[1]

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

[2]University of Queensland, Brisbane, Australia

AUTHOR ACCEPTED MANUSCRIPT

Please note that this paper is published in *Applied Ergonomics* and is reprinted here by permission of Elsevier. Please refer to this work using the complete citation shown below.

Wilson, M. D., 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

Understanding Fatigue in a Naval Submarine: Applying Biomathematical Models and Workload Measurement in an Intensive Longitudinal Design

Micah Kate Wilson[1], Timothy Ballard[2], Luke Strickland[1], Alex Boeing[1], Belinda Cham[1], Mark Griffin[1], Karina Jorritsma[1]

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

[2]University of Queensland, Brisbane, Australia

Author Note: This research was supported by funding from an Australian Defence Science and Technology Group Grant (MyIP:9079). Address correspondence to Micah Kate Wilson, Curtin University, 78 Murray Street, Perth, Western Australia 6000
The authors declare no competing interests. **Acknowledgements:** The authors would like to acknowledge and thank Dr Sam Huf for his continued and tireless efforts in supporting this research program.

Abstract

Fatigue is a critically important aspect of crew endurance in submarine operations, with continuously high fatigue being associated with increased risk of human error and long-term negative health ramifications. Submarines pose several unique challenges to fatigue mitigation, including requirements for continuous manning for long durations, a lack of access to critical environmental zeitgebers (stimuli pertinent to circadian physiology; e.g., natural sunlight), and work, rest and sleep occurring within an encapsulated environment. In this paper, we examine the factors that underlie fatigue in such a context with the aim of evaluating the predictive utility of a biomathematical model (BMM) of fatigue. Three experience sampling studies were conducted with submarine crews using a participant-led measurement protocol that included assessments of subjective sleepiness, workload (NASA-Task Load Index [TLX] and a bespoke underload-overload scale), and sleep. As expected, results indicated that predicting KSS with a BMM approach outperformed more conventional linear modelling approaches (e.g., time-of-day, sleep duration, time awake). Both the homeostatic and circadian components of the BMM were significantly associated with KSS and used as controls in the workload models. We found increased NASA-TLX workload was significantly associated with increased average KSS ratings at the between-person level. However, counter to expectations, the two workload measures were not found to have significant linear or quadratic relationship with fatigue at the within-person level. An important outcome of the research is that applied fatigue researchers should be extremely cautious applying conventional linear predictors when predicting fatigue. Practical implications for the submarine and related extreme work context are discussed. Important avenues for continued research are outlined, including directly estimating BMM parameters.

Keywords: fatigue, mental workload, extreme work environment, sleep, experience sampling methodology, biomathematical modelling, human performance

In extreme work environments, it is critical that, as the need arises, people are vigilant and alert to maintain maximal operational readiness and performance (Bell et al., 2016). Fatigue onset, typically defined as a physiological state of reduced mental or physical capability resulting from sleep deprivation, circadian processes, or task factors, is a critically important risk to human performance (Civil Aviation Safety Authority, 2014). Fatigue can cause impaired decision-making capabilities (Killgore et al., 2006), disturbances in socio-emotional processing (Dinges et al., 1997), and an increased risk of human error and performance deficits (Baulk et al., 2009; Dawson et al., 2011, 2017; Martin et al., 2019). Moreover, the effects of factors such as sleep loss, poor circadian entrainment, and high workloads on fatigue can accumulate over time, and if unaddressed, this can have long-term negative ramifications to physical and mental functioning (Dongen et al., 2003; McCauley et al., 2013).

Previous studies have investigated fatigue risk factors and sleep measurement protocols in safety-critical work domains such as rail transportation (Dorrian et al., 2011), surface navy crews (Grech et al., 2009), off-shore oil platforms (Riethmeister et al., 2019; Riethmeister et al., 2018), manufacturing (Baulk et al., 2009), simulated space missions (Flynn-Evans et al., 2020), and healthcare (Berastegui et al., 2020; Brzozowski et al., 2021; Karhula et al., 2013). In the current article, we investigate crew fatigue in a unique and important workplace environment: the military submarine. Maintaining crew endurance is crucial in the submarine, and degraded performance from fatigue can have serious safety implications in military environments (Comperatore et al., 2005; Miller et al., 2008; Shattuck et al., 2018). In some respects, the work challenges of submariners are analogous to challenges in other workplaces. For instance, much like shift-workers, the submariner sleep/wake cycle is often constrained to artificial watch-keeping structures (e.g., 6-on 6-off, Paul et al., 2010), and submariners must manage performance despite variability in work demand intensity causing fluctuations of underload and overload (Brasher et al., 2010). However, the submarine context poses several additional operational challenges relative to other safety-critical contexts.

A unique factor in the submarine context is that work activities, non-work activities and sleep are all performed within the constraints of an encapsulated environment (Brasher et al., 2010, 2012). It is not well understood how structuring sleep, respite, and work within such an environment may influence fatigue (Crain et al., 2018). Furthermore, the submarine environment limits exposure to sunlight, with uncertain impacts on circadian processes (Bass & Lazar, 2016). There are also tight limits on crew size, and submariners must maintain operational effectiveness over an extended period of time (e.g., weeks to months). As with any extreme work environment, significant forward planning is conducted in order to rotate crew members for respite.

Given the uniqueness of the submarine context, an important goal of the current research is to develop and test a fatigue measurement protocol, and to use this protocol to identify and test how factors such as workload are associated with fatigue risk. However, fatigue is affected by multiple non-linear, dynamic, physiological processes. To account for such components of fatigue, our analyses incorporate a biomathematical model (BMM) of fatigue that formally models the homeostatic and circadian processes underlying fatigue. Our study, therefore, aims to generate new insights derived from the unique context of submariners in addition to demonstrating a novel application of BMMs. Before introducing the current study, we introduce the theoretical underpinnings of fatigue and discuss potential factors affecting fatigue risk.

Causes of Mental Fatigue: Homeostatic, Circadian, and Workload Processes

Many key fatigue dynamics are captured by BMMs — a family of phenomenological models used to predict the neurobehavioral outcomes of fatigue (e.g., alertness, performance) on the basis of sleep/wake histories or work schedules (Dawson et al., 2017). BMMs are typically grounded in Borbély's (1982) seminal two-process model of sleep regulation. This model stipulates that neurobehavioral indicators of fatigue are modulated by the additive interaction of two biological processes: the homeostatic and the circadian. The homeostatic process, denoted by *S*, is responsible for the increase in fatigue during wake and the subsequent recovery from fatigue during sleep. The endogenous circadian process, denoted by *C*, reflects the effect of the circadian pacemaker on sleep propensity, and in turn mental fatigue. Figure 1 shows how predicted fatigue from a BMM varies across a 24-hour period based on variations in the *S* and *C* processes. Generally, differences between BMM implementations comprise the incremental addition and evaluation of functions presumed to reflect the underlying mechanisms of fatigue (Van Dongen, 2004). For instance, in addition to addition to processes *C* and *S*, the Three Process Model (TPM) — which has become ubiquitous across the literature — includes a sleep inertia process, *W*, that captures the alertness decrease that occurs immediately after waking (Åkerstedt et al., 2008; Åkerstedt & Folkard, 1997; Ingre et al., 2014).

BMMs are commonly applied for use as fatigue forecasting tools in aviation and defense sectors (Civil Aviation Safety Authority, 2014; Hursh et al., 2004). However, BMMs are rarely deployed in applied research settings, particularly those involving longitudinal experience sampling data. There are many theoretical, practical, and statistical advantages to applying BMMs as tools to better understand the dynamics of fatigue in the context of interest. For example, by using well validated group-level model configurations from prior research, BMM processes can be calculated from time-of-day and an individual's sleep history (Reifman & Gander, 2004). In turn, these BMM processes provide a more theoretically principled and effective method to capture fatigue dynamics compared to conventional methods, such as using "hours awake" as a linear predictor of fatigue. Below we provide further discussion of how fatigue prediction with circadian and homeostatic BMM processes can be more appropriate than conventional methods.

Figure 1.
Sensitivity plot of a Two Process Bio-Mathematical Model of Fatigue

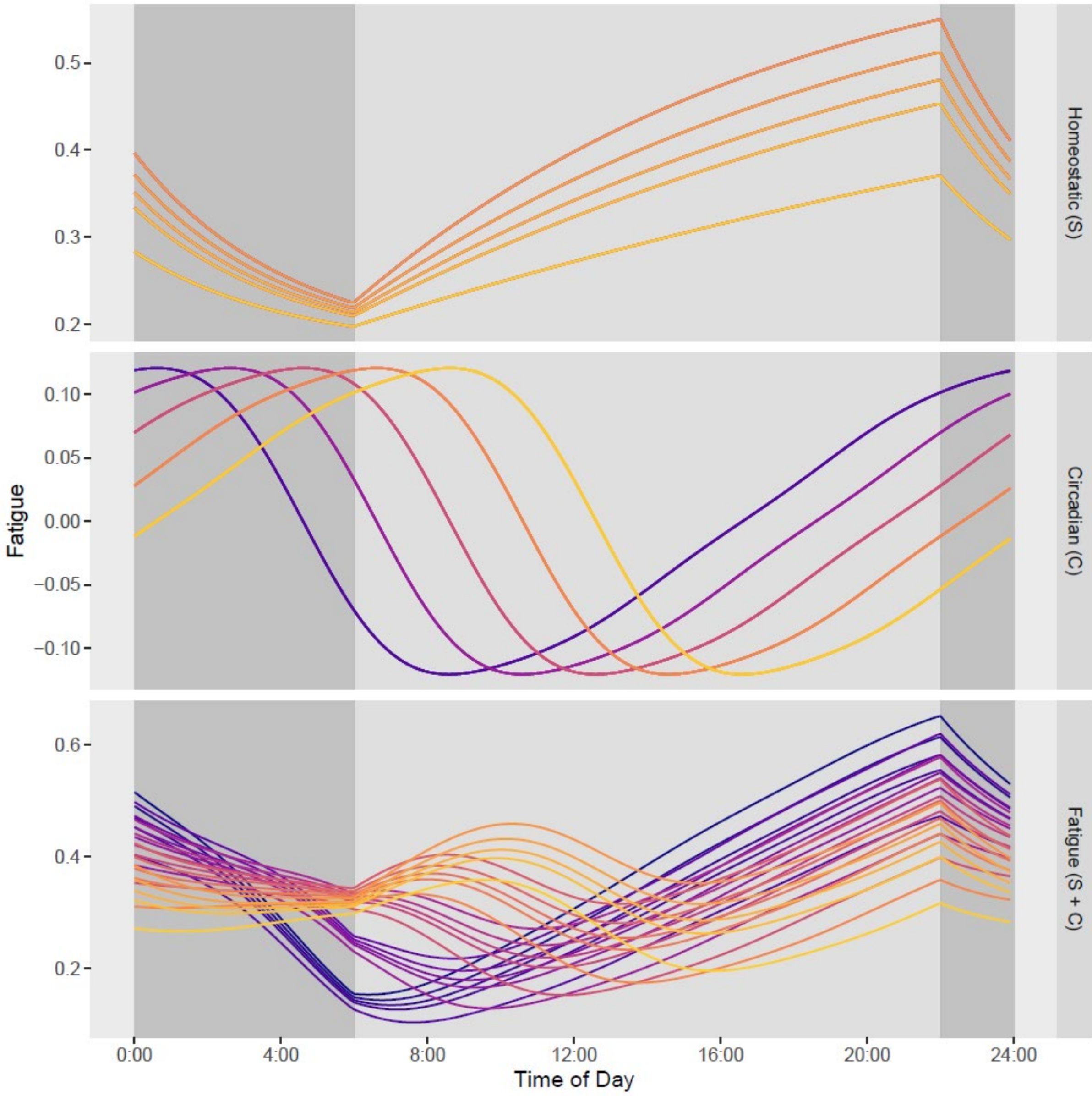


Figure note. The x axis represents a 24-hour day, with the dark gray plot regions indicating sleep and the light gray indicating wake. The top panel shows the homeostatic process with five variations of the τ_d parameter and the center panel shows the circadian process with five variations of the φ_{phase} parameter. The bottom panel shows the combinations of all unique *S* and *C* processes from the previous panels.

The influence of how fatigue functions in response to the time-of-day is challenging to model. It is well understood the circadian system's influence on fatigue is non-linear, with the dynamics typically modelled using a five-harmonic sinusoidal equation (i.e., the sum of five sine waves; Rajdev, 2013). This approach captures the variations in subjective fatigue caused by the circadian process, such as the afternoon dip, overall circadian peak, and overall low. An alternative approach from applied research involves using time-of-day as a fixed linear predictor of sleepiness (e.g., Berastegui et al., 2020; Grech et al., 2009; Riethmeister et al., 2019). However, such a linear assumption is inappropriate for circular number sets like time-of-day. For example, under a conventional linear approach, the difference in time-of-day from 11:00PM to 1:00AM would be modelled as 22 units, despite only being 2 hours apart. This is particularly problematic for research designs where observations are sampled sporadically across many times in a 24-hour day (e.g., environments that necessitate continuous staffing requirements).

Research using conventional linear modelling approaches has shown that time-in-bed and time awake can predict sleepiness and psychomotor vigilance performance, demonstrating the strong influence of homeostatic processes on fatigue (Berastegui et al., 2020). Additional predictors such as cumulative total sleep obtained over a given epoch (e.g., 24 hours) have also been found to capture historical influences (Bermudez et al., 2016; Riethmeister et al., 2019) — though there is little quantitative rationale that these variables accurately reflect the underlying biomechanics of fatigue. A limitation of modelling homeostatic processes with such variables is that linear models erroneously assume time-in-bed and time awake are independent predictors that each have a distinct linear mapping to fatigue. By contrast, BMMs model homeostatic fatigue impairment as a dynamic process, comprised of recovery during sleep and increased fatigue during wake using exponential functions with bounded upper and lower asymptotes. This approach implicitly captures effects of an individual's *entire* recorded sleep history on fatigue with a single variable (*S*).

This approach is particularly important in the submarine context, where split sleeping patterns and frequent naps are typical, thereby rendering linear predictors ineffective.

In summary, BMMs are potentially a powerful option available to researchers and enable the summarization of homeostatic and circadian processes in dynamic and non-linear terms. In doing so, BMMs provide a theoretically principled method of controlling for time-of-day and sleep history effects, enabling more rigorous examinations of the contributions of other factors which affect fatigue. Indeed, clearly there is scope for factors other than homeostatic and circadian influences to modulate fatigue, particularly in the context of a submarine. In particular, submariners are often engaged in demanding, and potentially draining work, that may lead to fatigue. In the following section, we describe how work factors and the intensity of work demands may influence fatigue

The Influence of Workload on Fatigue

In addition to sleep regulation processes, fatigue can also vary as a function of task demands such as mental workload or time pressure (Borghini et al., 2014; Grech et al., 2009; Hockey, 1997; Neal et al., 2014; Zohar et al., 2003), physical demands (Betts & Williams, 2010), and time-on-task (Lim et al., 2010). Mental workload characterizes an individual's assessment of how the demands of a task impose on their perceived information processing capacity (Wickens et al., 2015, p. 348). Both underload and overload must be managed to mitigate fatigue-related risks to operational effectiveness.

A growing body of research, typically using 'cohort designs', has shown consistently high workloads (i.e., *overload*) can increase operator fatigue (Baulk et al., 2007; Dorrian et al., 2011; Grech et al., 2009). While the mixed quality and methodological heterogeneity of this research has been criticized (see Studnek et al., 2018), there are strong theoretical justifications that workload may influence fatigue. Theories linking overload and fatigue have typically concentrated on compensatory control and arousal processes (Hockey, 1997).

Specifically, when task demands exceed an individual's capacity, they must engage compensatory processes to sustain task performance, requiring additional effort. Continuously engaging these compensatory processes is thought to deplete spare energetic reserves, leading to *active fatigue* which is associated with the core features of the subjective fatigue response (Desmond & Hancock, 2001; Saxby et al., 2013). In line with this, physiological evidence has shown that high workload is associated with short-term changes in cortisol secretion (Dahlgren et al., 2006; Schlotz et al., 2004). If experienced continuously, high workload can result in burnout, ongoing tension at work, and sleep disturbances (Melamed et al., 1999).

Although less extensively studied, critically low levels of workload (i.e., *underload*) can also negatively impact task performance and fatigue (Grech et al., 2009; Saxby et al., 2013; Young & Stanton, 2002). Underload situations are generally considered those in which an operator experiences a loss in task engagement and concentration owing to monotonous task demands that provide limited opportunities for the operator to exert active control (Desmond & Hancock, 2001; Saxby et al., 2013). Underload is thought to be associated with *passive fatigue,* whereby task demands are insufficient to raise arousal to the level required for sustaining alertness and engagement (Desmond & Hancock, 2001; Saxby et al., 2013).

Collectively, these studies and theoretical arguments give reason to expect the relationship between workload and fatigue is non-monotonic at the within-person or within-task level. Specifically, it may be characterized as a U-curve, whereby moderate levels of workload result in the least amount of fatigue and both low and high workload result in the most fatigue. To the extent that this is true, and that operators in the submarine environment face fluctuations between high and low workload situations, we would expect to see that the functional mapping between fatigue and workload follows this non-monotonical form.

Current Study

In the current research, three field studies were conducted in which teams of military submariners were measured multiple times per day for up to fourteen days using an intensive longitudinal experience sampling methodology. The broader program research was motivated by two overarching objectives: (1) to better understand the critical factors underlying mental fatigue in the submarine context, with particular focus on homeostatic factors and workload; and (2) to evaluate the applicability, predictive performance, and feasibility of using BMMs in a novel data context to aid inference. In this study, we compare two statistical approaches for modelling biological influences of fatigue: a 'conventional approach' implementing standard sleep-wake predictors (e.g., time-in-bed; time since awakening), to an approach that integrated information derived from a BMM. The outcome of this comparison will directly inform the most appropriate statistical controls to use for examining the role of workload. The comparison is also motivated by practical, methodological, and theoretical concerns. From a practical perspective, there is significant utility in evaluating the extent to which BMMs can capture and explain the dynamics of fatigue in a novel environment. However, to the extent that BMM approaches better account for the fatigue data in this specific domain, there may be broader methodological implications for best practice guidelines regarding modelling the processes underlying fatigue in field research.

Methods

Participants and Design

The data were collected as part of a broader program of research involving three at-sea submarine activities that took place during 2017 and 2018. The activities varied in operational tempo, with two being characterized as low to normal intensity, and the third activity being characterized as higher intensity. The study period of each activity (i.e., the duration of time we requested data to be captured) varied from 8 to 14 days. Participation in the trial was entirely voluntary, informed consent was gained from all participants, and participants were free to withdraw from the trial at any time. In total, 77 submariners representing a range of roles and organizational functions participated. Participants were at sea for the entire duration of the study period, following either a 6-on-6-off (n = 66) or 12-on-12-off watch schedule (n = 11). Work-related demographics are provided in Table 1 below which includes participants who attended pre-activity briefings but provided no other data.

Table 1.

Demographics associated with each of the three activities, median surveys completed per participant, and years of experience in their current role and in the navy overall.

Activity	*N*	Median Surveys [a]	Experience (Role)	Experience (Navy)
Low	22	40	3.62 years (SD = 3.06)	8.51 years (SD = 7.95)
Normal	39	39.5	3.69 years (SD = 3.98)	6.11 years (SD = 6.21)
High	16	28.5	2.08 years (SD = 1.10)	7.63 years (SD = 7.12)

[a] The median number of surveys completed per participant.

Measurement Protocol Structure

Each trial involved three phases: a briefing phase occurring pre-trial in which participants were familiarized with the purpose and requirements of the trial; an operational phase that involved the undertaking of the trial; and a debrief phase occurring post-trial that involved obtaining and providing feedback to participants. The measurement protocol reported here was developed iteratively over these three studies through interviews with operational personnel. It was designed to be minimally invasive to the participants and able to be carried out independently of researcher contact.

Pre-trial

Approximately one week prior to activity commencement, researchers conducted a pre-trial briefing to provide instructions for undertaking the trial, obtain informed participant consent, and communicate the purpose of the data collection. In this briefing, the study administrator issued participants a pack containing: a diary-survey booklet, pre- and post-trial surveys (not reported here), an actigraphy watch, and an ECG heart rate monitor (not reported here). The diary-survey booklet comprised a daily event log and daily surveys. The measures relevant to the current study are reported in measures section (below). The study administrator also provided instructions regarding how to wear the actigraphy watch, complete the surveys and event-logs, and the purpose of the study.

Operational phase

Participants were instructed to complete the pre-trial questionnaire and wear the actigraphy watch on the day prior to deployment. Throughout the duration of the activity, participants were instructed to record all activities performed in the event-log diary, which was recorded to 15-minute accuracy. Surveys were to be completed before and after every working period. Herein, we refer to these surveys as either ‘work’ or ‘non-work/rest’ surveys. At the conclusion of the trial (when the submarine returned to port), the experimenters met the submarine teams and participants were debriefed. The event logs and survey responses were digitized with a bespoke software application.

Measures

Sleep Measurement

Sleep was assessed with a combination of the sleep diaries and actigraphy data. Each participant was requested to complete a continuous diary of activities performed over each 24-hour period of the trial (e.g., “working”, “lying in bed”, “sleeping”). Throughout the duration of the activity, participants were instructed to record all activities performed in diary to 15-minute accuracy. Two previously validated actigraphy watches were used to measure sleep/wake times, the ‘Motion-watch 8’ by CamnTech and the ‘Motionlogger’ by Ambulatory Monitoring (Jean-Louis et al., 2001; Landry et al., 2015). Participants were instructed to wear the watch for the full duration of the trial.

Sleep diary records and activity records were cross-referenced to obtain estimates of time-in-bed. Specifically, two research officers inspected each sleep diary entry in a bespoke viewer alongside the actigraphy-derived sleep times. Diary-based sleep estimates that deviated from the actigraphy time by more than 10 minutes were set to the actigraphy time. In cases where a sleep period was detected by actigraphy, but no information was recorded in the diary *and* the participant was not rostered to work, the actigraphy-based sleep estimate was included.

Karolinska Sleepiness Scale (KSS)

We used the KSS to assess subjective sleepiness (i.e., subjective fatigue) due to its brevity and previous validation (Åkerstedt & Gillberg, 1990; Kaida et al., 2006). The single item asks, "*what is your current level of sleepiness?*" and is anchored from 1 ("*very alert"*) to 9 (*"very sleepy"*). The KSS was completed for each rest and work period, resulting in four observations per 24-hour period for individuals on a 6-on-6-off rotating watch schedule, and twice per 24-hour period for individuals on a 12-on-12-off rotating watch schedule.

Workload Measures

Two workload measures were used in the current study. Firstly, participants completed the NASA Task Load Index (Hart & Staveland, 1988; NASA-TLX). Participants did not perform item weighting, and thus 'raw TLX' scores were used (i.e., overall unweighted mean of items; Hart, 2006). Additionally, we developed a single-item measure of underload and overload. The item asked participants to rate "*the extent to which [they] felt underloaded or overloaded in [their] previous shift*", on a scale from 0–10. The scale was anchored from 0 ("*Underload: Where work requires concentration but there is little to do [unengaging].*") to 10 ("*Overload: Too much work to do, too little time [excessive work].*"). The item was developed to explore a scale which explicitly anchors underload (not just low workload), and to examine the convergent validity of a single-item workload measure for future uses where the full NASA-TLX is infeasible. The item was motivated by participant feedback from the low tempo activity, and therefore only included in the normal and high tempo activities. Both these workload measures were administered after each rostered shift, resulting in two observations per 24-hour period for individuals on a 6-on-6-off watch schedule, and once per 24-hour period for individuals on a 12-on-12-off watch schedule.

Biomathematical Modelling

Biomathematical modelling was conducted using the FIPS package (Wilson et al., 2020) for the R programming language (R Core Team, 2020). Model outputs (S, C and W) were generated with the Three Process Model (TPM) of Alertness (Akerstedt et al., 2004; Åkerstedt et al., 2008; Åkerstedt & Folkard, 1997) with the 'standard' model parameters as reported in Ingre et al. (2014). The TPM has undergone extensive validation studies and has been successfully used to model alertness in other safety-critical workplaces such as aviation (Folkard et al., 1999; Ingre et al., 2014). When using the standard model parameter settings, a linear transformation can be applied to scale the TPM model outputs direct to the KSS (Akerstedt et al., 2004; Åkerstedt & Folkard, 1997; Ingre et al., 2014).

The sleep times to generate the predicted alertness scores from the TPM were derived from the sleep times using the method reported above. That is, for each submariner, we input their sleep history into FIPS to generate predicted alertness across the activity. This procedure yielded estimates for the circadian (12 and 24 hour), homeostatic, and sleep inertia processes across the activity for each submariner.

Results

Compliance, Data Cleaning and Exclusion Criteria

The operational constraints of the context precluded the experimenters from being onboard the vessel during data collection, and so compliance was participant-led. This necessitated rigorous data integrity checks with several exclusion criteria. In total, 9 participants were excluded for not completing any of the protocol, and a further 26 participants were excluded from analyses due to: not completing diaries, actigraphy device not worn, not completing surveys, and/or not timestamping survey responses. A final sample size of 42 was achieved, comprising 39 crew members working on a 6-on-6-off watch schedule, and 3 working on a 12-on-12-off. A total of 1425 surveys (i.e., observations) were collected, with a median of 39 responses per participant. There was variability in the total duration participants remained in the study (as measured by the last survey response observed from the participant). Figure 1 plots the survival curve of observations across the three activities and shows all participants provided data responses for six days, with variability across activities in the dropout rates thereafter.

Figure 2.

Study dropout rate survival plot.

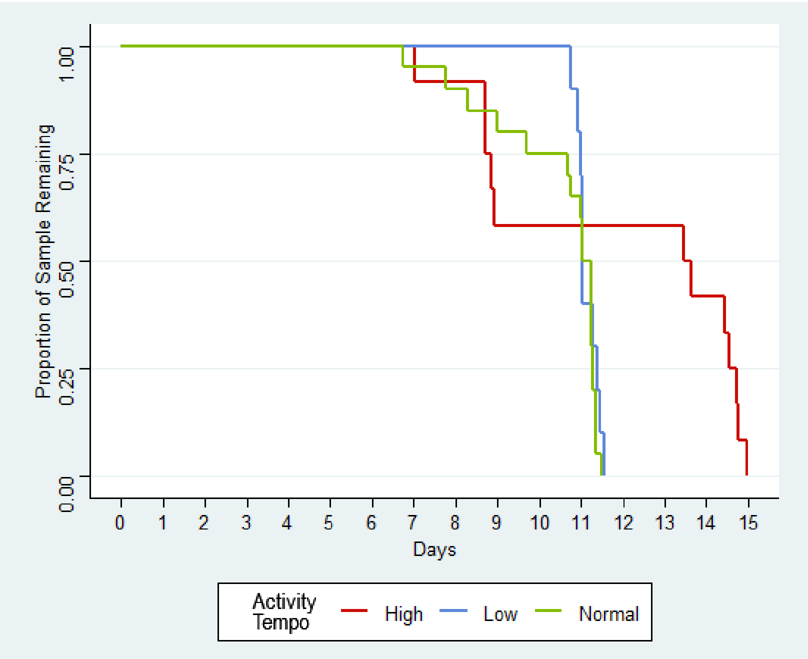


Figure note. Survival curves for sample proportion remaining for each activity across days. Survival was coded as the final survey observation datetime. The figure only includes participants retained for analysis. It is important to emphasise that large dips at the end of trials are expected, as they indicate the end of the data collection phase has occurred (i.e., 0% of sample remaining will inevidebly occur at the end of the study). If all participants finished data collection at the same time without any dropouts, the curve would be initally flat, with a straight vertical line down on the last day of the trial.

Hierarchical Fatigue Modelling

All reported statistical analyses and data wrangling was carried out in the R programming language and environment (R Core Team, 2020). Level 1 predictors (with the exception time-related variables) were grand mean centered unless otherwise specified. We applied hierarchical linear modelling (HLM) for hypothesis testing, using the lme4 R package (Bates et al., 2015). The study design was a cross-nested design as five participants completed both the normal and high operational tempo activities. Therefore, to satisfy independence assumptions given the longitudinal nature of the data, a participant-activity variable was used as the random effect. Significance was tested using Type II Wald chi-square tests (χ^2) implemented in the *car* package (Fox & Weisberg, 2019). Non-nested models were compared using the Akaike information criterion (AIC) – a well validated method of comparing the relative quality of statistical models (Akaike, 1998, p. 399; Bozdogan, 1987; Burnham & Anderson, 2004). AIC indexes goodness-of-fit and includes a penalty for overfitting, with a lower AIC indicating a better tradeoff between descriptive adequacy and model parsimony.

Fatigue HLM: Comparing Conventional Linear and BMM Approaches

The first objective of our fatigue HLM analysis is to contrast two modelling approaches―a BMM approach and a conventional linear predictor approach―to determine which best accounts for the dynamics of fatigue (i.e., variance in KSS scores). The data for the tested models was drawn from both work and non-work periods. The predictors from the superior model will be used as control variables in the subsequent workload models. We examined two HLMs: one for each modelling approach. The 'TPM computed values' (i.e., *C*, *S*, *W*) for each submariner were matched with the corresponding KSS rating by datetime (to the nearest minute). This enabled us to directly incorporate the TPM BMM within our hierarchical modelling framework (a similar approach to Van Dongen, 2004).

The conventional linear predictor model included three linear sleep-wake predictors: *time awake* (at measurement time in hours), *time-in-bed* (on previous sleep occasion in hours), and *work-status* (i.e., whether the survey was completed before or after a work period). KSS scores were significantly predicted by time-in-bed ($\chi^2 = 10.19$, $p < .001$) and work status ($\chi^2 = 73.81$, $p < .001$), but not time awake ($\chi^2 = 3.09$, $p = .079$). The BMM predictor model included the work-status predictor and the three time-varying process variables from the TPM: *S* (homeostatic), *C* (circadian), and *W* (sleep inertia). A random slope term for the circadian process was included to account for possible variation in circadian phase across individuals. All predictors were significant, with KSS scores being predicted by *S* ($\chi^2 = 39.29$, $p < .001$), *C* ($\chi^2 = 81.67$, $p < .001$), *W* ($\chi^2 = 5.35$, $p = .021$), and work status ($\chi^2 = 82.54$, $p < .001$). The model summary and parameter coefficients for both models can be inspected in Table 2. Critically, as predicted, both the AIC indices and R^2 estimates at the bottom of Table 2 demonstrate strong evidence favoring the BMM approach model relative to the conventional linear predictor model.

Table 2.
Model coefficient table for the two overall KSS score predictions models.

	Sleepiness (Fatigue) - Conventional		Sleepiness (Fatigue) - BMM	
Predictors	*Estimates*	*95% CI*	*Estimates*	*95% CI*
(Intercept)	4.79	4.40 — 5.19	6.09	5.44 — 6.74
Time-in-bed (hours)	-0.10	-0.16 — -0.04	—	—
Time awake (hours)	0.02	-0.00 — 0.04	—	—
Work status (after)	0.82	0.64 — 1.01	0.80	0.63 — 0.97
S (Homeostatic)	—	—	-0.17	-0.22 — -0.12
C (Circadian)	—	—	-0.23	-0.28 — -0.18
W (Sleep inertia)	—	—	-0.08	-0.14 — -0.01
Random Effects				
σ^2	2.22		2.00	
τ_{00}	0.88 $_{\text{Participant/Activity}}$		0.92 $_{\text{Participant/Activity}}$	
τ_{11}	—		0.01 $_{\text{Participant/Activity Circadian}}$	
ρ_{01}	—		0.06 $_{\text{Participant/Activity}}$	
ICC	0.28		0.32	
N	42 $_{\text{Participant/Activity}}$		42 $_{\text{Participant/Activity}}$	
Observations	1425		1425	
R^2 $_{\text{Marginal/Conditional}}$	0.073 / 0.336		0.135 / 0.411	
AIC	5314.90		5197.52	

Note. Marginal R^2 describes the proportion of variance attributable to the fixed effects alone, while the Conditional R^2 describes the variance explained by both the fixed and random effects. τ_{00} is the random intercept variance which indicates between-participant variation, while the residual variance σ^2 indicates the within-subject variance. τ_{11} is the random slope variance (i.e., circadian random slope) and ρ_{01} is the random slope-intercept correlation.

Workload-Specific Analyses

Next, we examine the relationship between mental workload (at both the within-person and between-person levels) and KSS scores. The two workload variables were the NASA-TLX scores (N = 704) and the overload-underload scale scores (N = 523). Note that the number of observations are reduced relative to the 'full data', as workload measures were not taken during rest periods. Separate models were fit for each workload measure. The between-person slope was obtained by using individual's within-person *mean* workload across the entire activity ($\bar{x}_i$); while the within-person slope was obtained by person-mean centering the time-varying workload ratings ($x_{ti} - \bar{x}_i$) (Curran & Bauer, 2011; Hamaker & Muthén, 2020). In both models below, KSS scores were predicted with between-person workload, quadratic orthogonal polynomials of the within-person workload, and the TPM[1] derived homeostatic and circadian process values. This model structure ensures any effects of workload must predict KSS scores over and above the homeostatic and circadian predictors. Full model coefficients are presented in Table 3.

First, we examine the NASA-TLX scores. Consistent with the BMM approach on the full data, KSS scores were significantly predicted by both *S* (χ^2 = 14.18, p < .001) and *C* (χ^2 = 43.61, p < .001). There was also a significant between-person effect of NASA-TLX scores (χ^2 = 4.71, p = .030), indicating that individuals with higher average workload (over an activity) reported increased sleepiness. However, at the within-person level, there was not a significant linear (χ^2 = 0.06, p = .813) or quadratic (χ^2 = 1.08, p = .299) relationship between NASA-TLX workload and KSS scores. For the underload-overload scale model, KSS scores were significantly predicted by both *S* (χ^2 = 5.93, p = .015) and *C* (χ^2 = 32.11, p < .001). However, there was not a significant between-person effect of underload-overload (χ^2 = 3.37, p = .066), and no significant within-person linear (χ^2 = 1.28, p = .257) or quadratic (χ^2 = 0.55, p = .459) relationships.

We also examined the association between the underload-overload and NASA-TLX ratings. Specifically, an HLM was fit specifying the NASA-TLX score as the dependent variable and underload-overload scale score as the predictor (and a random intercept term for participant). The model coefficients indicated that the underload-overload scale was significantly associated with NASA-TLX workload, β = 0.94, *SE* = .04, t = 23.21, p < .001. The overall model had an R^2 = 74.5% $_{\text{(Conditional)}}$ / 46% $_{\text{(Marginal)}}$, indicating a strong association between the NASA-TLX scores and underload-overload ratings (convergent validity).

1 The (W) parameter was not included as the KSS scores in the analyses below were taken at shift completion, where sleep inertia has ceased (Åkerstedt & Folkard, 1997).

Table 3.

Model coefficient table for the two overall workload predictor models.

	KSS (NASA-TLX)		KSS (Under/Overload)	
Predictors	*Estimates*	*95% CI*	*Estimates*	*95% CI*
(Intercept)	5.25	3.70 — 6.80	4.76	2.90 — 6.63
NASA-TLX (between)	0.35	0.03 — 0.67	—	—
NASA-TLX (within, linear)	-0.33	-3.06 — 2.40	—	—
NASA-TLX (within, quadratic)	1.54	-1.36 — 4.44	—	—
Homeostatic (s)	-0.15	-0.23 — -0.07	-0.11	-0.20 — -0.02
Circadian (c)	-0.32	-0.42 — -0.23	-0.30	-0.40 — -0.19
Under/Overload (between)	—	—	0.33	-0.02 — 0.68
Under/Overload (within, linear)	—	—	-1.58	-4.32 — 1.16
Under/Overload (within, quadratic)	—	—	1.12	-1.85 — 4.09
Random Effects				
σ^2	1.84		1.83	
τ_{00}	1.11 Participant/Activity		1.34 Participant/Activity	
τ_{11}	0.05 Participant/Activity Circadian		0.03 Participant/Activity Circadian	
ρ_{01}	0.07 Participant/Activity		0.08 Participant/Activity	
ICC	0.40		0.44	
N	42 Participant/Activity		32 Participant/Activity	
Observations	704		523	
R^2 Marginal/Conditional	0.147 / 0.489		0.132 / 0.512	
AIC	2575.510		1919.374	

Note. The two models should not be directly contrasted due to being fit to different subsets of the data. Marginal R^2 describes the proportion of variance attributable to the fixed effects alone, while the Conditional R^2 describes the variance explained by both the fixed and random effects. τ_{00} is the random intercept variance which indicates between-participant variation, while the residual variance σ^2 indicates the within-subject variance. τ_{11} is the random slope variance (i.e., circadian random slope) and ρ_{01} is the random slope-intercept correlation. CI is the 95% confidence interval.

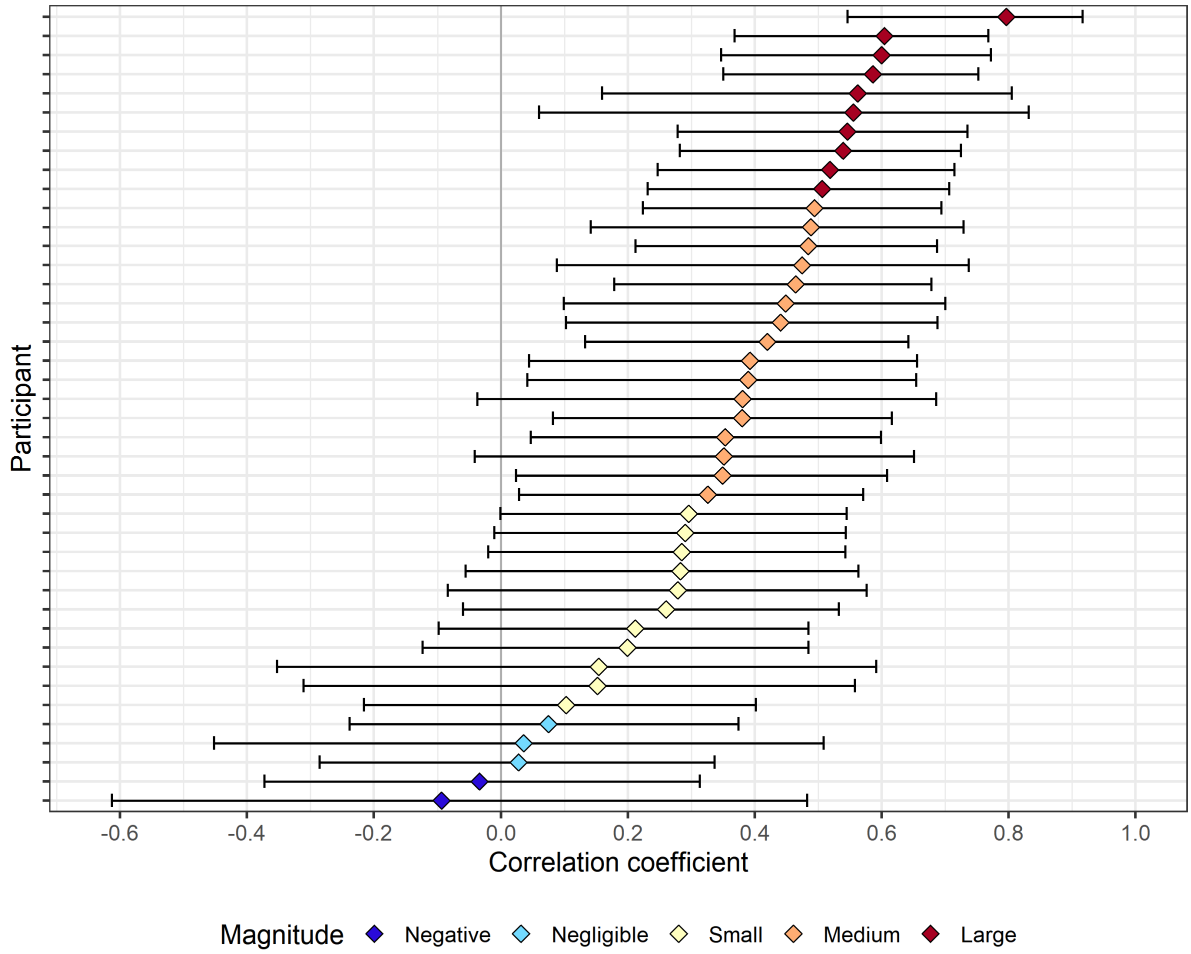
Evaluation of Model Predictive Performance

We found strong evidence that a regression model incorporating BMM predictors improved fit relative to conventional models. However, the BMM predictions themselves were not estimated from the observed data. That is, all participants' predictions were generated using the same set of governing parameters, but different sleep inputs (i.e., each submariner's sleep history). While conducting parameter estimation is beyond the scope of the current study, it is still important to evaluate the strength of association between the BMM predictions and the observed data (i.e., predictive performance), and to examine the extent of variability in this association across participants. Doing so helps to evaluate the utility of standard ‘off-the-shelf BMMs’ in the submarine context which also assume fixed group-level parameters. Further, the extent of variation across participants motivates future domain-specific parameter estimation efforts to improve predictive accuracy (Riedy et al., 2020).

To assess overall association between the predicted KSS scores from the Three Process Model (omitting *W* process) implemented in FIPS (Wilson, 2020) and the observed KSS ratings, we estimated the repeated measures correlation using the rmcorr R package (Bakdash & Marusich, 2020). We used Cohen’s (1988) guidelines to specify thresholds for negligible (0 – .1), small (.1 – .3), medium (.3 – .5) and large (.5 – 1) effect sizes, respectively. There was a positive significant repeated measures correlation of medium effect size between model predicted and observed KSS, $r_{(1382)} = 0.36$, $p < .001$, 95% CI [.31, .40]. In order to descriptively examine the variation in associative strength between individuals, a series of two-sided Pearson's product-moment correlations were performed between KSS scores and the model predictions for all participants independently (i.e., 42 correlations). Figure 3 shows the correlation coefficients across each participant categorized by effect size magnitude, and reveals the approximate distribution underlying the repeated measures correlation coefficient. In two cases (4.76%), the correlation coefficient was negative but negligible. For cases where a positive association was found: 38.10% (n = 16) had a medium association; 23.81% (n = 10) had a large association; 26.19% (n = 11) had a small association; and 7.14% (n = 3) had a negligible association. The implications of this are returned to in the discussion.

Figure 3.

Three process model and observed fatigue within-subject correlation plot.



Note. The plot shows the within-person correlations between the Three Process Model predictions and the observed Karolinska Sleepiness Scores (for each participant). Correlation coefficients are presented on the *x*-axis. Each point represents the correlation coefficient of one participant, with the fill color indicating magnitude according to Cohen's (1988) guidelines. Error bars show 95% parametric confidence intervals.

Discussion

In this study, we developed and tested a fatigue measurement protocol across three submarine activities, and implemented a BMM to better account for the dynamic and non-linear nature of the processes underlying fatigue. The application of a BMM in the unique context of the submarine environment (e.g., no natural lighting, continuous manning requirements) provided methodological and theoretical insights to other extreme-work environments (Bell et al., 2016). The results of the study can be summarized as follows. We found that a model incorporating the time varying BMM predictors (i.e., homeostatic, circadian and sleep inertia parameters) provided a substantially better account of the variation in submariner KSS scores than a model using linear temporal predictors (i.e., time-in-bed, wake duration). A repeated-measures correlation analysis revealed a medium-strength association between the observed and BMM predicted KSS scores; however, a descriptive individualized correlation analysis showed considerable variation in associative strength across participants. We found that NASA-TLX scores were associated with KSS scores at the between-person level, however within-person workload did not predict KSS scores over and above the BMM predictors.

BMM Approach: General Findings

An important contribution of the current study was the use of the TPM variables for predicting the KSS scores. In line with our expectations, incorporating a BMM that modelled the homeostatic, circadian, and sleep inertia dynamics provided superior fit compared to conventional linear predictor methods. In fact, within the conventional linear model, time awake was not found to be a significant predictor of sleepiness, a finding also reported by Berastegui et al. (2020). This reflects the broader inadequacies of conventional linear modelling approaches in accounting for the non-linear dynamics of fatigue. BMM frameworks provide a straightforward method of accounting for non-linear dynamics in a regression framework without requiring comprehensive non-linear modelling techniques (c.f., Mollicone et al., 2010).

Practically speaking, a core goal of this study was to determine the feasibility of BMMs in the submariner context. In addition to the findings above, the medium-strength repeated-measures correlation (between the BMM predictions and observed fatigue) indicates substantial support for the use of BMMs in a submarine context for general scheduling purposes, such as comparing prospective rosters (see also, Flynn-Evans et al., 2020).

However, there was non-trivial variation in predictive accuracy across participants which does preclude drawing inferences or recommending BMM application at the individual level in this context (e.g., individualized fatigue management). The choice to use fixed BMM governing parameters in the current study is likely to account for some of this variability. Ideally, model parameters should be context-tailored or individualized by conducting parameter estimation, either at the group level or hierarchically, which would substantially improve prediction accuracy. Further, this would offer insights into between-person variability in the biomechanical processes, and would be essential for personalized real-time fatigue evaluation (Liu et al., 2017). However, parameter estimation of the TPM using the current dataset was infeasible in this study due to data constraints. An important next step would be to collect sufficient data to estimate the BMM parameters directly based on the fatigue ratings, ideally at an individual level. Finally, it should be noted that the superior predictive success of the TPM indicates circadian processes are important to submariner fatigue, suggesting that optimizing circadian patterns with roster design or lighting technologies may be promising avenues of fatigue intervention.

Work Demands and Workload

Submariners reported significantly higher KSS scores after a work-shift, even when accounting for the homeostatic process (*S*) in the model. This is consistent with research showing that fatigue increases as a function of time-on-task in functionally similar ways to homeostatic and circadian related fatigue (Khosroshahi, 2019; Lim et al., 2010; Veksler & Gunzelmann, 2018). In terms of workload, submariners who reported higher levels of NASA-TLX workload, averaged across an activity (i.e., between-person level), also reported higher KSS scores. In contrast to our expectations, there was no significant linear or quadratic within-person relationship between KSS ratings and either the NASA-TLX or the underload-overload scale (controlling for homeostatic and circadian processes). While the betweensubjects workload effect is largely consistent with the results of previous research (Baulk et al., 2007; Dorrian et al., 2011; Grech et al., 2009; Karhula et al., 2013), the underpinning theoretical accounts of fatigue-workload interactions have typically focused on within-person explanations (e.g., within-task compensatory control mechanism; Desmond & Hancock, 2001). There are several possible explanations of our pattern of results.

A likely possibility is that overload situations may have a serial dependency with fatigue, whereby overload induced stress may have deleterious effects on subsequent sleep and recovery (i.e., knock-on effect; Crain et al., 2018). Indeed, this mechanism may underlie our observed between-person workload effect. Such a mechanism may manifest at the within-person level via a cross-lagged effect of workload on fatigue (e.g., past workload on current fatigue). In our data, the temporal irregularities precluded using discrete-time dynamic regression models that can tractably evaluate cross-lagged effects. With sufficiently large sample size, continuous-time dynamic structural equation modelling may be a viable alternative in future research (Driver & Voelkle, 2018).

A further explanation is that, at the within-person level, workload may only influence an individual's *sensitivity* to fatigue (Baulk et al., 2007). Thus, underload or overload may additively increase fatigue *only* when homeostatic pressure is high. Indeed, this is similar to how Peng et al., (2018) incorporate task demands in their BMM-based performance model. There are at least two explanations underlying the lack of a within-person *quadratic* effect of workload. First, given arousal and control processes can fluctuate relatively fast (Veksler & Gunzelmann, 2018), it is conceivable that our fatigue and retrospective workload ratings were insensitive to more momentary fluctuations. Secondly, it is possible the findings of prior research (e.g., Grech et al., 2009) may fail to replicate if within and between person variance are partitioned (Curran & Bauer, 2011), and homeostatic or circadian processes controlled for. Admittedly, the unique context and sampling methodology of this study, in conjunction with heterogeneity of other fatigue-workload research (Studnek et al., 2018), does prevent straightforward reconciliation of our specific findings with past literature.

An important practical outcome was the finding that the underload-overload scale demonstrated convergent validity with the NASA-TLX. Throughout developing our measurement protocol, submariners indicated desire for an expeditious single-item workload measure that explicitly included underload. Though further validation may be required, we believe the scale could be useful in other experience sampling studies of work environments that are characterized by underload and overload fluctuations, but where routinely completing the full NASA-TLX is not logistically feasible for participants.

Limitations, Frontiers for Practice, and Conclusions

The fact this study was conducted in an operational submarine context did constrain the comprehensiveness of the measurement protocol and the generalizability of our findings. For instance, in terms of measurement, despite 18 crew members being provided actigraphy watches capable of performing the psychomotor vigilance task (PVT), only 6 routinely completed the task (far too few observations for robust statistical analysis). Similarly, work-related constraints precluded measurement of workload during a shift. As workload measures could only be captured at the end of a shift, a limitation is they represented a retrospective assessment of the workload from the previous six hours. This limitation cannot be easily rectified in our current design out of safety concerns (i.e., submariners cannot be expected undertake PVTs mid operations), but we acknowledge that more temporally sensitive metrics of within-shift workload and workload variation would provide better insights to within-subject fatigue and workload mechanisms. Additionally, although our intensive experience sampling design afforded high ecological validity, the exact mechanics underlying the relationship between fatigue and workload are likely to differ under controlled laboratory experimentation. Continued research investigating how this relationship changes in response to situational context (e.g., work domain) and time scale (e.g., days to weeks) is crucially important for advancing the development of fatigue models.

A practical direction for future research will be to move beyond identifying and predicting the occurrence of fatigue, but to evaluate the mitigatory effectiveness of proactive work design interventions (e.g., structured recovery activities; protected sleep times; circadian-optimized lighting). This is particularly important in any 24/7 continuous operation environment where the work system may provide limited scope for recovery activities and quality sleep hygiene.

In closing, the current study has identified several important factors underlying fatigue processes in extreme work environments, and provided practical and methodological directions for future research. The results of the current study showed that work-related factors influence fatigue, over and above the effects of circadian and homeostatic factors. Beyond domain-specific findings, this study has direct implications for both practitioners and researchers examining fatigue. Importantly, we have demonstrated and discussed the clear limitations of unilateral application of conventional linear approaches to fatigue research, and provided new methodological directions leveraging techniques from BMM literature. Further, our results generally support the use of BMMs in novel contexts, which has direct implications for practitioners aiming to apply these methods in extreme work contexts with analogous workplace constraints to the military submarine.

References

Akaike, H. (1998). Prediction and Entropy. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 387–410). Springer. https://doi.org/10.1007/978-1-4612-1694-0_30

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

Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, *52*(1–2), 29–37. https://doi.org/10.3109/00207459008994241

Åkerstedt, T., Ingre, M., Kecklund, G., Folkard, S., & Axelsson, J. (2008). Accounting for Partial Sleep Deprivation and Cumulative Sleepiness in the Three-Process Model of Alertness Regulation. *Chronobiology International*, *25*(2–3), 309–319. https://doi.org/10.1080/07420520802110613

Khosroshahi B. E., (2019). *A Unified Model of Fatigue in a Cognitive Architecture* (Doctoral dissertation, Drexel University, Philadelphia, United States). https://search.proquest.com/docview/2328377233/abstract/C3799B09924944F1PQ/1

Bass, J., & Lazar, M. A. (2016). Circadian time signatures of fitness and disease. *Science*, *354*(6315), 994. https://doi.org/10.1126/science.aah4965

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baulk, S. D., Fletcher, A., Kandelaars, K. J., Dawson, D., & Roach, G. D. (2009a). A field study of sleep and fatigue in a regular rotating 12-h shift system. *Applied Ergonomics*, *40*(4), 694–698. https://doi.org/10.1016/j.apergo.2008.06.003

Baulk, Stuart D., Kandelaars, K. J., Lamond, N., Roach, G. D., Dawson, D., & Fletcher, A. (2007). Does variation in workload affect fatigue in a regular 12-hour shift system? *Sleep and Biological Rhythms*, *5*(1), 74–77. https://doi.org/10.1111/j.1479-8425.2006.00249.x

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

Berastegui, P., Jaspar, M., Ghuysen, A., & Nyssen, A.-S. (2020). Fatigue-related risk perception among emergency physicians working extended shifts. *Applied Ergonomics*, *82*, 102914. https://doi.org/10.1016/j.apergo.2019.102914

Bermudez, E. B., Klerman, E. B., Czeisler, C. A., Cohen, D. A., Wyatt, J. K., & Phillips, A. J. K. (2016). Prediction of Vigilant Attention and Cognitive Performance Using Self-Reported Alertness, Circadian Phase, Hours since Awakening, and Accumulated Sleep Loss. *PLOS ONE*, *11*(3), e0151770. https://doi.org/10.1371/journal.pone.0151770

Betts, J. A., & Williams, C. (2010). Short-Term Recovery from Prolonged Exercise. *Sports Medicine*, *40*(11), 941–959. https://doi.org/10.2165/11536900-000000000-00000

Borbély, A. A. (1982). A two process model of sleep regulation. *Human Neurobiology*, *1*(3), 195–204.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370. https://doi.org/10.1007/BF02294361

Brasher, K. S., Sparshott, K. F., Weir, A. B. C., Day, A. J., & Bridger, R. S. (2012). Two year follow-up study of stressors and occupational stress in submariners. *Occupational Medicine*, *62*(7), 563–565. https://doi.org/10.1093/occmed/kqs104

Brasher, Kate S., Dew, A. B. C., Kilminster, S. G., & Bridger, R. S. (2010). Occupational stress in submariners: The impact of isolated and confined work on psychological well-being. *Ergonomics*, *53*(3), 305–313. https://doi.org/10.1080/00140130903067763

Brzozowski, S. L., Cho, H., Arsenault Knudsen, É. N., & Steege, L. M. (2021). Predicting nurse fatigue from measures of work demands. *Applied Ergonomics*, *92*, 103337. https://doi.org/10.1016/j.apergo.2020.103337

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

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

Comperatore, C. A., Rivera, P. K., & Kingsley, L. (2005). Enduring the shipboard stressor complex: A systems approach. *Aviation, Space, and Environmental Medicine*, *76*(6 Suppl), B108-118.

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

Curran, P. J., & Bauer, D. J. (2011). The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change. *Annual Review of Psychology*, *62*(1), 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Dahlgren, A., Kecklund, G., & Akerstedt, T. (2006). Overtime work and its effects on sleep, sleepiness, cortisol and blood pressure in an experimental field study. *Scandinavian Journal of Work, Environment & Health*, *32*(4), 318–327. https://doi.org/10.5271/sjweh.1016

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

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

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., & Pack, A. I. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*, *20*(4), 267–277.

Dongen, V., P.a, H., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The Cumulative Cost of Additional Wakefulness: Dose-Response Effects on Neurobehavioral Functions and Sleep Physiology From Chronic Sleep Restriction and Total Sleep Deprivation. *Sleep*, *26*(2), 117–126. https://doi.org/10.1093/sleep/26.2.117

Dorrian, J., Baulk, S. D., & Dawson, D. (2011). Work hours, workload, sleep and fatigue in Australian Rail Industry employees. *Applied Ergonomics*, *42*(2), 202–209. https://doi.org/10.1016/j.apergo.2010.06.009

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

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

Folkard, S., Åkerstedt, T., Macdonald, I., Tucker, P., & Spencer, M. B. (1999). Beyond the Three-Process Model of Alertness: Estimating Phase, Time on Shift, and Successive Night Effects. *Journal of Biological Rhythms*, *14*(6), 579–589. https://doi.org/10.1177/074873099129000911

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Grech, M. R., Neal, A., Yeo, G., Humphreys, M., & Smith, S. (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), 231–242. https://doi.org/10.1037/a0014952

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, *25*(3), 365–379. https://doi.org/10.1037/met0000239

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*, 904–908.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Hockey, R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1–3), 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4

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.

Ingre, M., Van Leeuwen, W., Klemets, T., Ullvetter, C., Hough, S., Kecklund, G., Karlsson, D., & Åkerstedt, T. (2014). Validating and Extending the Three Process Model of Alertness in Airline Operations. *PLoS ONE*, *9*(10), e108679. https://doi.org/10.1371/journal.pone.0108679

Jean-Louis, G., Kripke, D. F., Mason, W. J., Elliott, J. A., & Youngstedt, S. D. (2001). Sleep estimation from wrist movement quantified by different actigraphic modalities. *Journal of Neuroscience Methods*, *105*(2), 185–191. https://doi.org/10.1016/S0165-0270(00)00364-2

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, *117*(7), 1574–1581. https://doi.org/10.1016/j.clinph.2006.03.011

Karhula, K., Härmä, M., Sallinen, M., Hublin, C., Virkkala, J., Kivimäki, M., Vahtera, J., & Puttonen, S. (2013). Association of job strain with working hours, shift-dependent perceived workload, sleepiness and recovery. *Ergonomics*, *56*(11), 1640–1651. https://doi.org/10.1080/00140139.2013.837514

Killgore, W. D. S., Balkin, T. J., & Wesensten, N. J. (2006). Impaired decision making following 49 h of sleep deprivation. *Journal of Sleep Research*, *15*(1), 7–13. https://doi.org/10.1111/j.1365-2869.2006.00487.x

Landry, G. J., Best, J. R., & Liu-Ambrose, T. (2015). Measuring sleep quality in older adults: A comparison using subjective and objective methods. *Frontiers in Aging Neuroscience*, *7*. https://doi.org/10.3389/fnagi.2015.00166

Lim, J., Wu, W., Wang, J., Detre, J. A., Dinges, D. F., & Rao, H. (2010). Imaging Brain Fatigue from Sustained Mental Workload: An ASL Perfusion Study of the Time-On-Task Effect. *NeuroImage*, *49*(4), 3426–3435. https://doi.org/10.1016/j.neuroimage.2009.11.020

Martin, K., Périard, J., Rattray, B., & Pyne, D. B. (2019). Physiological Factors Which Influence Cognitive Performance in Military Personnel. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872081984175. https://doi.org/10.1177/0018720819841757

McCauley, P., Kalachev, L. V., Mollicone, D. J., Banks, S., Dinges, D. F., & Van Dongen, H. P. A. (2013). Dynamic Circadian Modulation in a Biomathematical Model for the Effects of Sleep and Sleep Loss on Waking Neurobehavioral Performance. *Sleep*, *36*(12), 1987–1997. https://doi.org/10.5665/sleep.3246

Melamed, S., Ugarten, U., Shirom, A., Kahana, L., Lerman, Y., & Froom, P. (1999). Chronic burnout, somatic arousal and elevated salivary cortisol levels. *Journal of Psychosomatic Research*, *46*(6), 591–598. https://doi.org/10.1016/S0022-3999(99)00007-0

Miller, N. L., Matsangas, P., & Shattuck, L. G. (2008). Fatigue and its effect on performance in military environments. *Performance Under Stress*, 231–249.

Mollicone, D. J., Van Dongen, H. P. A., Rogers, N. L., Banks, S., & Dinges, D. F. (2010). Time of Day Effects on Neurobehavioral Performance During Chronic Sleep Restriction. *Aviation, Space, and Environmental Medicine*, *81*(8), 735–744. https://doi.org/10.3357/ASEM.2756.2010

Neal, A., Hannah, S., Sanderson, P., Bolland, S., Mooij, M., & Murphy, S. (2014). Development and Validation of a Multilevel Model for Predicting Workload Under Routine and Nonroutine Conditions in an Air Traffic Management Center. *Human Factors*, *56*(2), 287–305. https://doi.org/10.1177/0018720813491283

Paul, M. A., Hursh, S. R., & Miller, J. C. (2010). *Alternative Submarine Watch Systems* (Technical Report TR 2010-001). DRDC Toronto.

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

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

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

Reifman, J., & Gander, P. (2004). Commentary on the three-process model of alertness and broader modeling issues. *Aviation, Space, and Environmental Medicine*, *75*(3), A84–A88.

Riedy, S. M., Fekedulegn, D., Andrew, M., Vila, B., Dawson, D., & Violanti, J. (2020). Generalizability of a biomathematical model of fatigue's sleep predictions. *Chronobiology International*, *37*(4), 564–572. https://doi.org/10.1080/07420528.2020.1746798

Riethmeister, V., Matthews, R. W., Dawson, D., de Boer, M. R., Brouwer, S., & Bültmann, U. (2019). Time-of-day and days-on-shift predict increased fatigue over two-week offshore day-shifts. *Applied Ergonomics*, *78*, 157–163. https://doi.org/10.1016/j.apergo.2019.02.010

Riethmeister, Vanessa, Bültmann, U., Gordijn, M., Brouwer, S., & de Boer, M. (2018). Investigating daily fatigue scores during two-week offshore day shifts. *Applied Ergonomics*, *71*, 87–94. https://doi.org/10.1016/j.apergo.2018.04.008

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied*, *19*(4), 287–300. https://doi.org/10.1037/a0034386

Schlotz, W., Hellhammer, J., Schulz, P., & Stone, A. A. (2004). Perceived Work Overload and Chronic Worrying Predict Weekend–Weekday Differences in the Cortisol Awakening Response: *Psychosomatic Medicine*, *66*(2), 207–214. https://doi.org/10.1097/01.psy.0000116715.78238.56

Shattuck, N. L., Matsangas, P., & Dahlman, A. S. (2018). Sleep and Fatigue Issues in Military Operations. In E. Vermetten, A. Germain, & T. C. Neylan (Eds.), *Sleep and Combat-Related Post Traumatic Stress Disorder* (pp. 69–76). Springer New York. https://doi.org/10.1007/978-1-4939-7148-0_7

Studnek, J. R., Infinger, A. E., Renn, M. L., Weiss, P. M., Condle, J. P., Flickinger, K. L., Kroemer, A. J., Curtis, B. R., Xun, X., Divecha, A. A., Coppler, P. J., Bizhanova, Z., Sequeira, D. J., Lang, E., Higgins, J. S., & Patterson, P. D. (2018). Effect of Task Load Interventions on Fatigue in Emergency Medical Services Personnel and Other Shift Workers: A Systematic Review. *Prehospital Emergency Care*, *22*(sup1), 81–88. https://doi.org/10.1080/10903127.2017.1384874

Van Dongen, H. (2004). Comparison of mathematical model predictions to experimental data of fatigue and performance. *Aviation, Space, and Environmental Medicine*, *75*(3), A15–A36.

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

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering Psychology & Human Performance*. Taylor & Francis.

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

Zohar, D., Tzischinski, O., & Epstein, R. (2003). Effects of energy availability on immediate and delayed emotional reactions to work events. *The Journal of Applied Psychology*, *88*(6), 1082–1093. https://doi.org/10.1037/0021-9010.88.6.1082