Can We Predict Cognitive Performance Decrements Due to Sleep Loss and the Recuperative Effects of Caffeine?

Sridhar Ramakrishnan,[1] Srinivas Laxminarayan,[1] Nancy J. Wesensten,[2] Thomas J. Balkin,[2] and Jaques Reifman[1]

[1]Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD 21702
UNITED STATES

[2]Department of Behavioral Biology, Walter Reed Army Institute of Research, Silver Spring, MD 20910
UNITED STATES

sramakrishnan@bhsai.org; slaxminarayan@bhsai.org; nancy.j.wesensten.civ@mail.mil; thomas.j.balkin.civ@mail.mil; jaques.reifman.civ@mail.mil

ABSTRACT

Warfighters are often subjected to challenging sleep/wake schedules that hinder their cognitive performance. Countermeasures, such as timely short naps and caffeine, are often used to mitigate the effects of sleep loss on performance. However, the timing, duration, and dosage of these countermeasures should be optimal (or near optimal) to maintain high levels of performance at the desired times. To this end, mathematical models that can accurately predict the detrimental effects of sleep loss and the restorative effects of different dosages of caffeine on performance could be of great utility. Here, we present a mathematical model that predicts individualized cognitive performance for subjects exposed to the continuum of sleep loss (from no sleep to partial sleep) while considering the recuperative effects of caffeine. In particular, we developed and validated both group-average and individual-specific models on performance data obtained from four different studies. Results from the first two studies showed that a group-average model developed on one study could accurately predict the temporal dynamics of both total and partial sleep loss in another study, with >75% of the predictions falling within 2 standard errors of the measured data. The results also showed that, on average, individual-specific models provided ~30% improved prediction accuracy compared with the group-average models. Importantly, we showed that once the model had been customized to an individual under total sleep loss, it could be directly applied to predict the same individual's performance under partial sleep loss, and vice versa. Results from the third and fourth studies showed that the group-average model that accounts for caffeine effects on performance can provide up to 90% improved accuracy in capturing the effects of a range of caffeine doses (50-300 mg) over a model that does not account for caffeine. Here, individual-specific models provided up to 43% improved accuracy in predicting caffeine effects on performance compared with caffeine-free models. Taken together, the proposed model is the first mathematical formulation that allows for the prediction of the effects of any form of sleep loss and the restorative effects of different dosages of caffeine on a specific individual's performance. When used as a decision-aid tool, this model provides the means to maximize Warfighter cognitive performance, resulting in peak alertness and prolonged alertness at the desired times.

1.0 INTRODUCTION

During combat operations, military personnel are often subjected to challenging sleep/wake schedules, ranging from a few days of no sleep to many days of fewer than 6 h of sleep per night. This naturally hinders their

cognitive performance, leading to an increased risk of errors and accidents [1]. Caffeine doses and timely short naps can help mitigate the effects of sleep loss on performance. However, the timing, duration, and dosage of these countermeasures need to be optimal (or near optimal) to maintain high levels of performance at the desired times. Hence, mathematical models that can accurately predict the effects of sleep loss and the restorative effects of different durations of naps and dosages of caffeine on performance would be highly desirable. Such models could be used as a mission-planning tool to help commanders quantitatively assess the state of Warfighters in real time and to determine the optimal countermeasure schedule for each Warfighter to sustain performance for the entire duration of the mission.

Recognizing this need, the U.S. Department of Defense organized a Fatigue and Performance Modeling Workshop, held on June 13-14, 2002, in Seattle, WA, with the objectives of determining the state of the art in biomathematical models of performance and, more importantly, identifying research gaps [2]. The key findings from this Workshop were as follows:

- All models were based on Borbély's seminal two-process model of sleep regulation [3], which provides an elegant phenomenological representation of the temporal dynamics of sleep propensity as a function of sleep and wake schedules. While these models performed well when applied to humans exposed to total sleep deprivation (TSD) conditions, they were less accurate in predicting the effects of partial sleep loss [or chronic sleep restriction (CSR)] because they did not account for the effects of prior sleep debt [4].
- All models were "group-average" models, representing the average performance of a group of individuals. In other words, they did not account for the substantial inter-individual variability that exists with regard to response to sleep loss.
- None of the models accounted for the recuperative effects of caffeine on performance.

Here, we address the above-mentioned research gaps by developing a mathematical model that can *1*) predict the effects of both TSD and CSR conditions, *2)* be customized to an individual to provide individual-specific predictions of performance during sleep loss, and *3*) predict the effects of caffeine on performance. In the sections below, we present the methods and results as well as a brief discussion on each of these capabilities.

2.0 UNIFIED MODEL OF PERFORMANCE FOR BOTH TOTAL AND PARTIAL SLEEP LOSS

Here, we developed and validated a unified model of performance (UMP) that can account for effects of both total and partial sleep loss in a single (i.e., unified) mathematical modeling framework. Specifically, we used performance data from one study, comprising of four different CSR conditions, to develop the model and validated its predictions on performance data from another study, involving a TSD and a CSR condition.

2.1 Methods: Unified Model of Performance

2.1.1 Two-process Model

The two-process model postulates that the temporal pattern of performance can be represented as the additive interaction of two processes [5]. The first, process *S*, represents the homeostatic influence on performance wherein the homeostat increases during wakefulness and decreases during sleep such that these increases/decreases operate within fixed upper and lower asymptotes that are independent of prior sleep debt. The second process is the endogenous circadian rhythm, process *C*, which is independent of the sleep/wake

history and represents a self-sustaining oscillator with a 24-h period [5]. Mathematically, performance $P(t)$ at time instant t can be expressed as follows:

$$P(t) = S(t) + \kappa C(t), \tag{1}$$

$$\frac{dS(t)}{dt} = \begin{cases} [U - S(t)]/\tau_w \text{during wakefulness} \\ [L - S(t)]/\tau_s \text{ during sleep,} \end{cases} \tag{2}$$

$$C(t) = \sum_{i=1}^{5} a_i \sin\left[i \frac{2\pi}{24}(t + \phi) \right], \tag{3}$$

where κ represents the circadian amplitude, U and L denote the upper and lower asymptotes of process S, respectively, τ_w and τ_s denote the time constants of the increasing and decreasing sleep pressure during wakefulness and sleep, respectively, a_i, $i = 1, \ldots, 5$, represent the amplitude of the five harmonics ($a_1 = 0.97$, $a_2 = 0.22$, $a_3 = 0.07$, $a_4 = 0.03$, and $a_5 = 0.001$) governing process C, and ϕ denotes the circadian phase.

2.1.2 The Unified Model

The UMP was developed as an extension of the two-process model [6]. In the UMP, process S is dependent on prior sleep debt such that the capacity to recover during sleep varies inversely with extant sleep debt. Specifically, the UMP modulates the lower asymptote L of process S as a function of the sleep debt resulting from prior sleep/wake history such that the most recent sleep loss exerts the greatest effect, with the sleep loss influence decreasing with increasing temporal distance. Mathematically, we represented the lower asymptote $L(t)$ and sleep debt $Debt(t)$ as follows:

$$L(t) = U \times Debt(t), \tag{4}$$

$$\frac{dDebt(t)}{dt} = [Loss(t) - Debt(t)]/\tau_{LA}, \tag{5}$$

where τ_{LA} denotes the time constant of the exponential decay of the effect of sleep history on performance, and $Loss(t)$ assumes a value of 1 during wakefulness and –2 during sleep [6]. Further, to better account for the effects of extended sleep under well-rested conditions, we imposed a minimum level of −0.11 on $Debt(t)$.

The UMP thus consists of eight parameters: *1*) U, the upper asymptote of homeostatic process S; *2*) τ_w, the time constant of increasing homeostatic pressure during wake time; *3*) τ_s, the time constant of decreasing homeostatic pressure during sleep; *4*) S_0, the initial state value for process S; *5*) κ, the amplitude of circadian process C; *6*) ϕ, the circadian phase; *7*) τ_{LA}, the time constant accounting for the exponential rise and fall of sleep debt (via modulation of the lower asymptote L) as a function of sleep/wake history; and *8*) L_0, the initial state value of L. The first six parameters originate from the original two-process model, whereas the last two parameters account for the effects of sleep debt.

2.1.3 Study Data

We obtained psychomotor vigilance task (PVT) data from two previously published sleep studies to develop and validate the UMP. The PVT is a visual vigilance task in which subjects press a button in response to a visual stimulus that is presented as a random interval (2-10 s) schedule over a 10-min period, resulting in ~100 stimulus-response pairs [7, 8]. In a PVT session, a time (initially set to "000") is displayed in the center of a computer screen, and subjects are instructed to press a response key as soon as the time display begins to scroll.

The subject's response stops the timer, displays the reaction time for ~0.5 s, and then initiates the next trial. The subject's response times (RTs) are automatically recorded, which are then used to compute two commonly used statistics: *1*) lapses (number of RTs >500 ms) and *2*) mean RT (average of all RTs >100 ms and <3000 ms).

The first study (labeled *Study A*) involved 57 healthy adults (age: 24-62 yr, mean: 38 yr) who underwent 7 consecutive nights of 3, 5, 7, or 9 h of time in bed (TIB) during the CSR phase followed by 3 consecutive nights of 8-h TIB for recovery in a controlled laboratory study [9]. A 10-min PVT was administered 4 times/day (09:00, 12:00, 15:00, and 21:00 hours). Subjects in the 3- and 5-h TIB study conditions performed additional PVT sessions (at 00:00 hours for both study conditions and again at 02:00 hours for the 3-h TIB study condition) during their additional time awake.

The second study (labeled *Study B*) involved 19 healthy adults (age: 18-39 yr, mean: 28 yr) who underwent 2 sleep-loss conditions (crossover design) separated by 2-4 wk: *1*) 64-h TSD and *2*) CSR consisting of 7 consecutive nights of 3-h TIB [10]. Both conditions were preceded by 7 in-laboratory nights with 10-h TIB, and followed by 3 nights with 8-h TIB (recovery). During the entire wake period of TSD/CSR and recovery, 10-min PVTs were administered every 2 h.

In both studies, subjects slept for ~8 h each night for at least 3 days prior to the study.

2.1.4 Fitting and Prediction

We used group-average performance data from *Study A*, involving four different CSR conditions, to estimate the model parameters by minimizing the sum of squared errors between the group-average data and the model outputs. We then used this model to predict the group-average performance data in *Study B*, involving both TSD and CSR conditions. This allowed us to investigate whether the model based on sleep/wake conditions in one study could simultaneously capture the effects of TSD, CSR, and variation in sleep/wake history on performance in another study.

2.1.5 Goodness of Fits

To assess the goodness of fit and predictions, we calculated the root mean squared errors (RMSEs) between UMP outputs $P(t)$ and group-average performance data for each study condition. For the predictions, we also quantified the likelihood that the predictions came from the same distribution as the group-average data. Specifically, we computed the percentage of model predicted points that lie within a 95% confidence interval (CI) of the group-average data, wherein we used the standard errors of the data to compute CIs (95% CI ≈ 2 standard errors) [11]. Thus, higher percent values indicate greater likelihood and, hence, better predictions.

2.2 Results: Unified Model of Performance

We used group-average PVT lapse data from *Study A* to estimate the UMP model parameters. We then used this model to predict the group-average lapse data in *Study B*.

2.2.1 Model Fits on *Study A*

We obtained the following parameter estimates (standard error) by fitting the UMP on *Study A* lapse data: U = 18.35 (0.73) lapses, τ_w = 40.00 (3.19) h, τ_s = 2.11 (0.11) h, S_0 = 0.50 (0.66) lapses, κ = 3.26 (0.26) lapses, ϕ = 2.31 (0.26) h, τ_{LA} = 7.00 (1.67) days, and L_0 = 0.00 (0.00) lapses. Figure 1 shows the corresponding fits on data from each of the four study conditions in *Study A*. The UMP was able to capture the dose-dependent effect of

TIB on performance during CSR, yielding RMSEs ranging from 1.20 lapses (for the 7-h TIB condition) to 3.26 lapses (for the 3-h TIB condition). It also successfully captured the slower recovery in the 3-h TIB condition compared with the recovery times in the other study conditions.

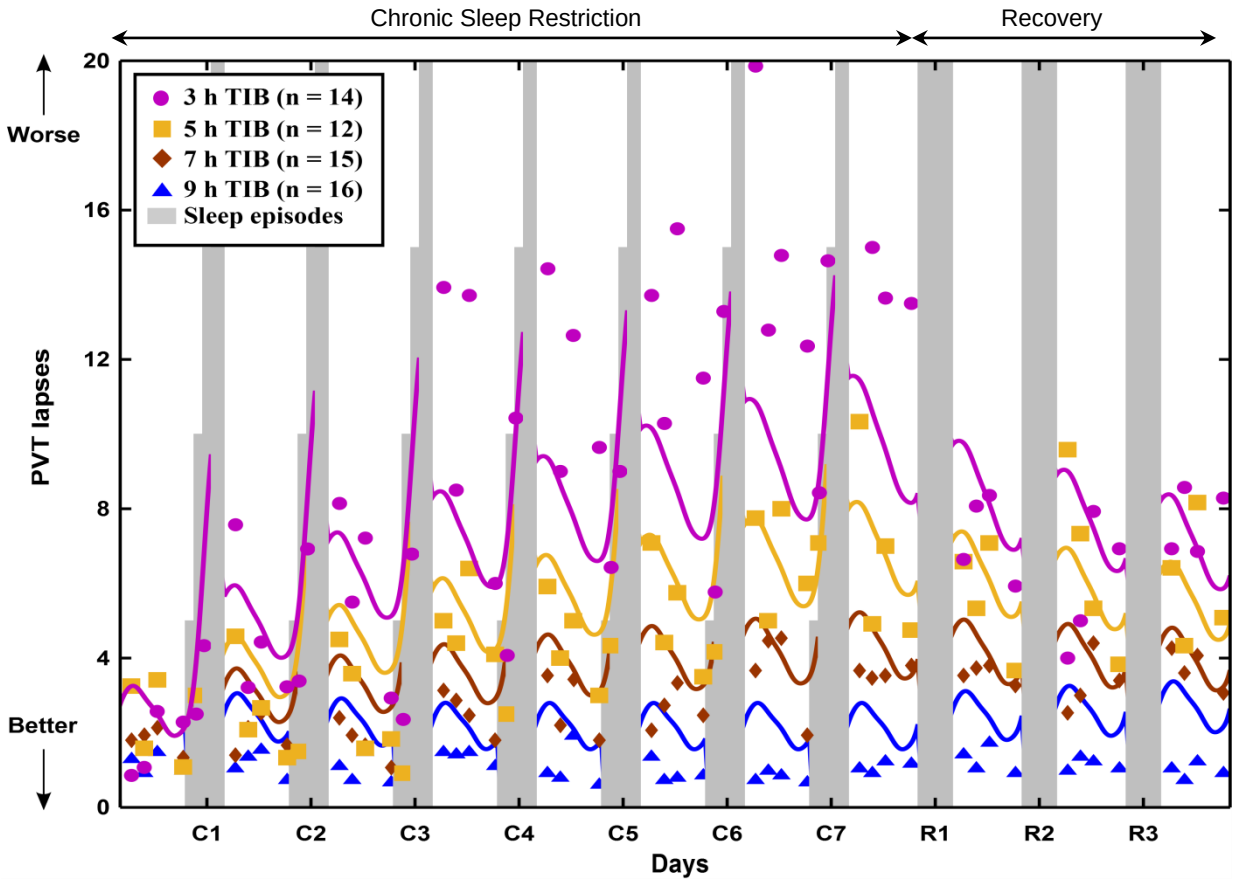

Figure 1: Group-average psychomotor vigilance task (PVT) lapse data and unified model of performance (UMP) fits on 3-, 5-, 7-, and 9-h time in bed (TIB) groups for 7 nights of chronic sleep restriction (CSR; C1-C7) followed by three 8-h TIB recovery nights (R1-R3) in *Study A* [9].

2.2.2 Model Predictions on *Study B*

Using the model fitted on *Study A* data, we predicted performance on both TSD and CSR conditions in *Study B*. Figure 2 shows group-average lapse data and UMP predictions during the TSD/CSR and recovery phases for *Study B*. Except for the PVT sessions between T1 and T2 (TSD condition) and the first session (i.e., the first data point) immediately after sleep between C4 and C7 (CSR condition), the UMP accurately predicted the effects of sleep loss during both TSD and CSR (and the recovery phases after them), yielding overall RMSEs of 2.46 and 2.39 lapses, respectively. Also, for TSD and CSR, 80% and 76% of the model predictions fell within the 95% CIs of the measured data, respectively. As observed in the data, the UMP also predicted a faster recovery after 64-h TSD compared with CSR of 7 nights of 3-h TIB.

2.3 Discussion: Unified Model of Performance

The UMP extends the classical two-process model of sleep regulation by accounting for the effects of a known sleep/wake history on performance, and it bridges the continuum from short periods of acute TSD to longer periods of sleep restriction (hence, the term "unified"). Importantly, it does not require an artificial distinction between sleep deprivation, restricted sleep, and recovery sleep, that is, it simultaneously represents both TSD and CSR performance data. And, because it modifies only those aspects of the two-process model that relate to recovery of the homeostatic process during sleep, it reduces to Borbély's original model as TIB approaches zero hours (i.e., under conditions of acute TSD).

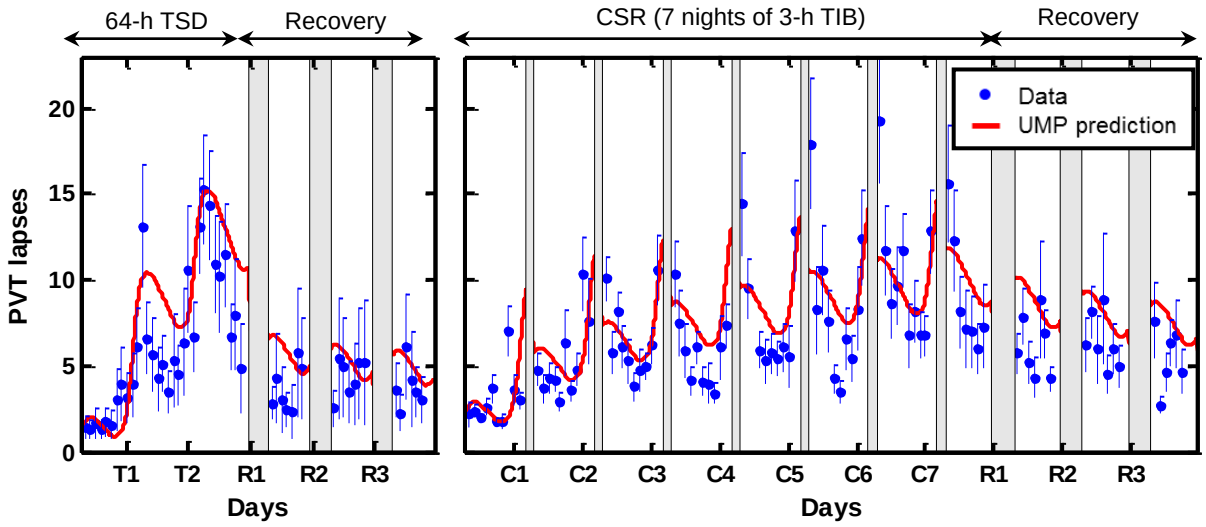

Figure 2: Group-average lapse data and standard errors and UMP predictions during total sleep deprivation (TSD; T1-T2), CSR (C1-C7), and recovery (R1-R3) phases in *Study B* [10]. Gray-shaded vertical bars represent sleep episodes.

The results from the present analyses indicate that the UMP is capable of accurately capturing the dose-dependent effects of TIB during CSR and subsequent recovery on PVT performance. Furthermore, it accurately predicted group-average performance under both 64-h TSD and 3-h TIB CSR conditions of another study, with >75% of predictions falling within 2 standard errors of the measured data.

One of the practical utilities of the UMP is to be able to simulate the impact of any given sleep/wake schedule on performance without the need to perform additional experimental studies. Accordingly, it could serve as a useful tool to design and optimize laboratory sleep-study protocols and serve as a key component of fatigue management systems.

3.0 CAN THE UMP PREDICT AN INDIVIDUAL'S TRAIT-LIKE RESPONSE TO SLEEP LOSS?

Historically, biomathematical performance prediction models have focused on predicting group-average performance impairment. However, laboratory studies have repeatedly shown that inter-individual differences account for nearly 70% of the total variance in performance among a group of individuals [12]. Importantly, studies have also shown that these differences are trait like and stable [10, 12]. That is, an individual's response to one sleep-loss condition is positively correlated to the same individual's response to a subsequent, different sleep-loss condition.

The purpose of the present work was threefold: *1*) to develop individual-specific UMP models, *2*) to determine the extent to which a UMP developed for an individual under one sleep-loss condition could predict the same individual's response under a different sleep-loss condition, and *3*) to determine the extent to which individual-specific models increase prediction accuracy over group-average models.

We used the PVT mean RT data from *Study B* (described in 2.1.3 Study Data) for these analyses [10]. We excluded four subjects from our analyses: one subject was excluded due to missing data and the others were excluded due to significant differences in RT distributions of their baseline sessions (first day of TSD/CSR) between the TSD and CSR conditions.

3.1 Methods: Modeling Individual Differences During Sleep Loss

3.1.1 Individual-specific Models

To obtain individual-specific models of performance, we fitted the UMP to each subject's PVT performance data obtained from each of the two sleep-loss conditions (TSD and CSR), resulting in two sets of model parameters for each of the 15 subjects. Because the UMP output was insensitive to 3 trait parameters, τ_w, τ_s, and τ_{LA} (i.e., the model output did not change appreciably with changes in these parameters), we fixed them to physiologically meaningful values, e.g., τ_w = 10 h, τ_s = 2 h, and τ_{LA} = 7 days [6, 13]. We thus estimated only five individual-specific model parameters (U, κ, S_0, L_0, and ϕ) for each subject.

Using the UMP parameters developed from the TSD data, we computed the individual-specific model fits (P_i) under TSD and the corresponding cross-condition predictions (P_{Xi}) under CSR, and vice versa.

3.1.2 Group-average Prediction Models

For comparison purposes, we also developed two group-average prediction models for every subject, one for each sleep-loss condition. To obtain a group-average model for predicting subject i, we excluded performance data from the i-th subject when averaging the data used for model fitting. We then used these models to compute cross-condition predictions ($\overline{P}_{Xi}$), generating two sets of predictions for each subject i, one set for each condition, based on the two group-average models.

3.1.3 Comparing Model Fits and Predictions

To compare accuracies of the individual-specific model fits (P_i), group-average model predictions ($\overline{P}_{Xi}$), and individual-specific model predictions (P_{Xi}), we calculated the RMSE between each subject's performance data and the corresponding model fits or model predictions.

3.2 Results: Modeling Individual Differences During Sleep Loss

3.2.1 Individual-specific Model Fits vs. Predictions

Table 1 lists the five UMP parameters for each subject under each condition (entries within parentheses correspond to the TSD condition) obtained by fitting the model to subjects' mean RT data measured under the two conditions. Also listed are the means and standard deviations for each parameter. We observed that for every subject, each of the parameters was similar across the two conditions, with the average difference being <15% for all parameters except for ϕ, which exhibited an average absolute difference of ≤0.5 h. (Because ϕ exhibits a 24-h periodicity, it is improper to compute a percent difference for it.)

Using the individual-specific UMP parameters obtained from TSD data, we computed the corresponding fits to TSD performance data and predictions of performance under the CSR condition. We performed similar computations for the CSR data. Figure 3 shows the model fits (P_i) and cross-condition predictions (P_{Xi}) for three different subjects (Subjects #3, #5, and #12), who showed different patterns of response to sleep loss. Individual-specific fits (blue solid lines) accurately captured the within- and across-day performance variations under both TSD and CSR. To a slightly lesser extent, the cross-condition predictions (red dashed lines) were also accurate. (The RMSEs of the predictions were no greater than 5 ms compared with those of the fits.) The fits and cross-condition predictions were accurate during the recovery days as well, except for the first recovery day after TSD for Subject #3 and for the second recovery day after CSR for Subject #12; in these instances, the subjects appeared to recover faster than predicted by the models.

Table 1: Individual-specific parameters of the unified model of performance for each of the 15 subjects based on psychomotor vigilance task (PVT) data (mean response time statistic) under chronic sleep restriction and total sleep deprivation (entries within parentheses) conditions in *Study B* [10].

Subject	U (ms)		κ (ms)		S_0 (ms)		L_0 (ms)		ϕ (h)	
1	270	(296)	31	(37)	207	(200)	207	(180)	1.2	(1.0)
2	410	(385)	33	(30)	200	(200)	200	(180)	3.0	(2.4)
3	289	(281)	34	(32)	200	(200)	180	(180)	2.4	(4.5)
4	302	(283)	50	(48)	200	(200)	180	(180)	1.0	(1.5)
5	248	(238)	17	(15)	200	(209)	190	(181)	1.0	(4.1)
6	333	(289)	31	(44)	200	(200)	200	(180)	2.8	(5.7)
7	253	(254)	15	(18)	200	(200)	180	(180)	2.8	(2.1)
8	250	(236)	15	(15)	236	(214)	180	(214)	1.0	(1.0)
9	292	(271)	16	(30)	224	(200)	221	(200)	1.0	(1.5)
10	228	(230)	15	(15)	200	(204)	180	(180)	1.0	(1.0)
11	245	(237)	18	(20)	200	(200)	180	(180)	1.0	(1.5)
12	369	(395)	37	(50)	260	(222)	225	(188)	1.6	(1.0)
13	311	(366)	44	(50)	200	(200)	200	(180)	1.4	(1.0)
14	259	(253)	15	(15)	200	(200)	180	(184)	1.0	(1.0)
15	348	(385)	49	(50)	200	(200)	180	(188)	1.0	(1.0)
Mean	294	(293)	28	(31)	208	(203)	192	(185)	1.5	(2.0)
SD	52	(60)	13	(14)	18	(7)	16	(10)	0.8	(1.5)

L_0, lower homeostatic asymptote at time zero; S_0, homeostatic state at time zero; SD, standard deviation; U, upper asymptote of the homeostatic process; ϕ, circadian phase; κ, circadian amplitude

3.2.2 Individual-specific vs. Group-average Model Predictions

Figure 4 shows the RMSEs of the individual-specific fits (P_i), cross-condition predictions based on individual-specific models (P_{Xi}), and cross-condition predictions based on group-average models ($\overline{P}_{Xi}$) for both the TSD and CSR conditions. As expected, for all subjects, RMSEs of P_i were smaller than those of P_{Xi}. However, the differences between them were not statistically significant (for both conditions, mean difference over the 15 subjects = 4 ms, SD = 4 ms), implying that the individual-specific cross-condition model predictions were

almost as good as the fits. In fact, on average, P_{Xi} yielded only 9% and 14% higher RMSEs than P_i under TSD and CSR conditions, respectively. In contrast, RMSEs of both P_i and P_{Xi} were significantly lower than those of $\overline{P}_{Xi}$ (P <0.05 by a Wilcoxon paired, two-sided, signed-rank test [14]), with P_{Xi} yielding, on average, 35% and 29% lower RMSEs than $\overline{P}_{Xi}$ under TSD and CSR conditions, respectively.

3.3 Discussion: Modeling Individual Differences During Sleep Loss

A key finding of this work is that the UMP captures an individual's unique trait-like response to sleep loss. In other words, a model developed for an individual on one sleep-loss condition can be used to predict performance of the same individual under another sleep-loss condition. The individual-specific cross-condition model predictions (P_{Xi}) yielded only marginally higher RMSEs (up to 14%) than the corresponding fits (P_i). However, they yielded up to 35% lower RMSEs than the corresponding cross-condition group-average model predictions ($\overline{P}_{Xi}$), thereby advocating the use of individual-specific models over group-average models for predicting individuals' performance under either of the two sleep-loss conditions. An expanded treatment of these findings was reported in Ref. 15.

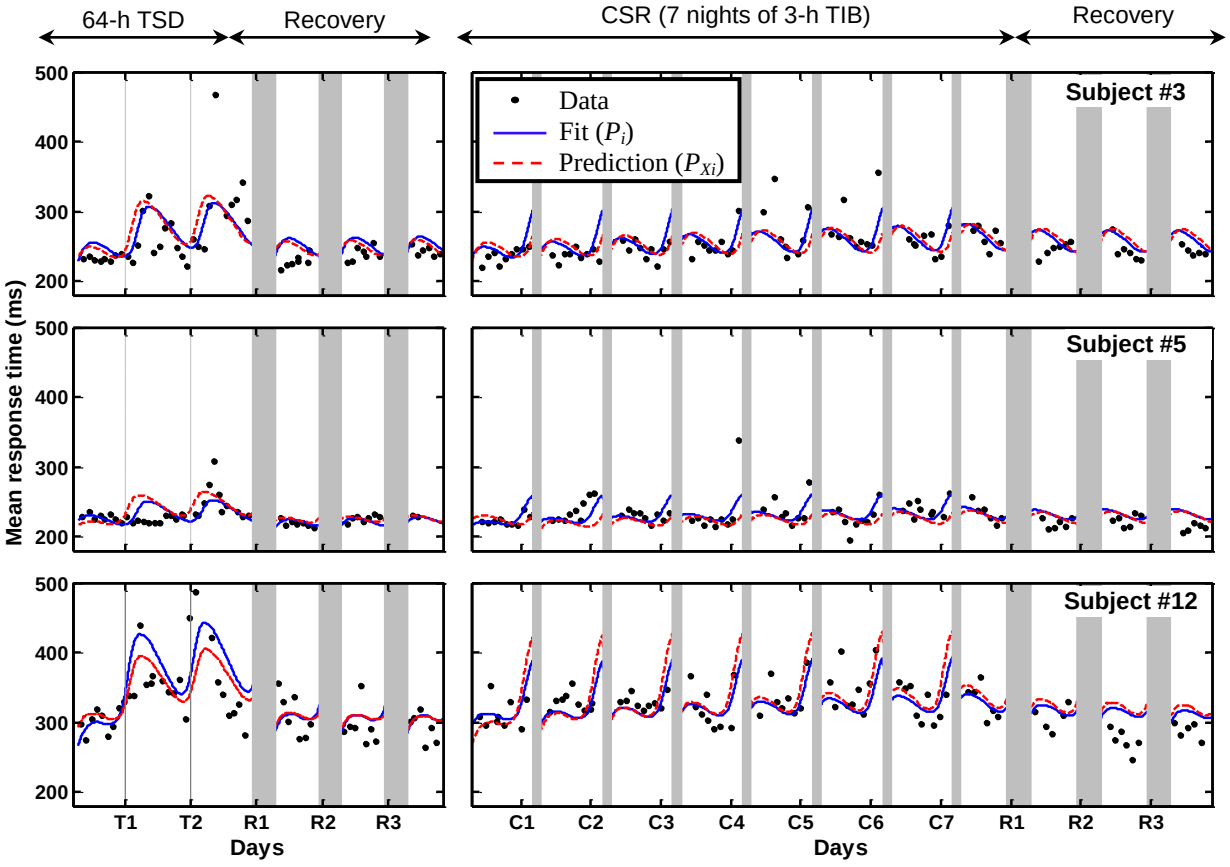

Figure 3: Individual-specific fits (P_i) and cross-condition predictions (P_{Xi}) of PVT performance (mean response time statistic) using the UMP for three different subjects challenged to TSD and CSR conditions in *Study B* [10]. For each sleep-loss condition, the solid blue lines represent the fits and the dashed red lines represent the cross-condition predictions based on models obtained by fitting on data from the other condition. Shaded regions represent the sleep episodes.

In this work, we applied the UMP model on PVT performance data for individuals who underwent total and partial sleep loss under laboratory conditions. However, the model may also be applicable to individuals in operational environments using available computational platforms. This would require information about the individual's sleep-wake history and performance data, which together would serve as inputs to the UMP model running on a computational platform [tablet, personal computer (PC), or a smartphone] to customize the model and make predictions. For example, an individual's sleep-wake history could be continuously inferred via wrist-worn actigraphy and streamed to a computer. Similarly, brief, periodic PVT performance tests [16] could be performed on a computer running the UMP model [17].

4.0 DOSE-DEPENDENT MODEL OF CAFFEINE EFFECTS ON PERFORMANCE DURING SLEEP LOSS

Caffeine is the most widely consumed stimulant to counter sleep-loss effects. While the pharmacokinetics of caffeine in the body are well understood, its alertness-restoring effects are still not well characterized. In fact, mathematical models capable of predicting the effects of varying doses of caffeine on objective measures of cognitive performance are not available.

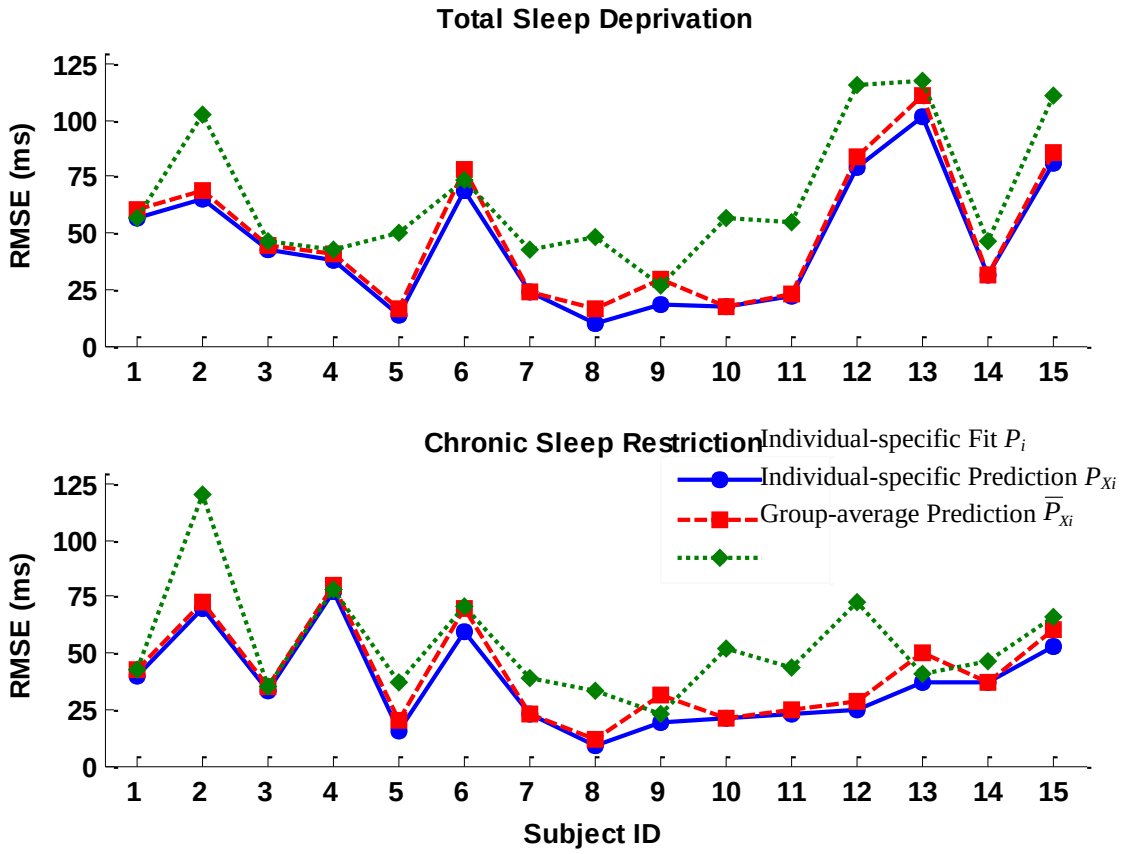

Figure 4: Root mean-squared errors (RMSEs) of model fits (P_i), cross-condition individual-specific model predictions (P_{Xi}), and cross-condition group-average model predictions ($\overline{P}_{Xi}$) of PVT performance (based on mean response time statistic) during TSD and CSR for each of the 15 subjects in *Study B* [10].

Here, we developed and validated a biomathematical model that extends the model for sleep loss by accounting for the performance-restoring effects of varying doses of caffeine. Specifically, we used PVT lapse data from one study (involving repeated administration of caffeine doses of 0, 50, 100, and 200 mg) to develop the model and validated the model's predictions on lapse data from another study (involving repeated administration of caffeine doses of 0, 100, 200, and 300 mg) at both group-average and individual levels.

4.1 Methods: Modeling Dose-dependent Effects of Caffeine

4.1.1 Model of Dose-dependent Effects of Caffeine

We hypothesized that caffeine has a multiplicative effect on performance during sleep loss. In other words, the performance impairment estimate [$P_c(t)$] at time instant t of a sleep-deprived individual after caffeine intake can be formulated as follows:

$$P_c(t) = P(t) \times g_{PD}(t), \tag{6}$$

where $P(t)$ represents the individual's performance in the absence of caffeine (referred to as caffeine-free performance henceforth) at time awake t and $g_{PD}(t)$ represents the caffeine effect factor, with $0 \leq g_{PD}(t) \leq 1$, where 1 corresponds to pharmacodynamic (PD) effects in the absence of caffeine, i.e., the most impaired performance, and 0 corresponds to the maximal PD effect on cognitive performance, i.e., complete restoration with no impairment [18]. Hence, with this multiplicative model, performance impairment levels decrease after caffeine intake and, eventually, as caffeine is cleared, return to the levels that would be observed if caffeine had not been administered. Here, we used the two-process model [Eq. (1)] to characterize the caffeine-free performance $P(t)$.

To model the caffeine effect factor (g_{PD}), we used the one-compartment pharmacokinetic (PK) model of caffeine [19, 20] within the Hill equation [21] that relates PK and PD, and expressed g_{PD} of caffeine dose D administered at time t_0 as follows:

$$g_{PD}(t, D) = \left\{1 + M_D \exp\left[-k_D (t - t_0)\right]\right\}^{-1} \quad \text{for } t \geq t_0, \tag{7}$$

where M_D and k_D represent the amplitude factor and elimination rate parameters of the caffeine model, respectively, which depend on caffeine dose D. To account for the dose-dependent effects of caffeine, we used a linear and exponential model to capture the effects of dose on M_D and k_D, respectively, in accordance with prior results from studies on PK and PD of caffeine under TSD [22-24]. Accordingly, we expressed M_D and k_D as follows:

$$M_D = M_0 \times D \quad \text{and} \quad k_D = k_0 \exp(-zD), \tag{8}$$

where M_0 denotes the amplitude slope, k_0 denotes the basal elimination rate, and z denotes the decay constant. In what follows, we refer to M_0, k_0, and z as the dose-dependent caffeine model parameters.

To extend the g_{PD} model for repeated caffeine doses, we modified Eqs. (7-8) so that at the beginning of each dose the amplitude factor M_D and elimination rate k_D accounted for the extant plasma caffeine concentration, which was computed using the standard one-compartment PK model [25, 26]. Accordingly, the PD effect after j doses of caffeine of strengths $D_1, D_2, \ldots, D_j$ administered at times $t_1, t_2, \ldots, t_j$, respectively, can be expressed as follows:

$$g_{PD}(t, D_j) = \begin{cases} 1 & \text{for } t < t_1 \\ \left\{1 + M_{Dj} \exp\left[-k_{Dj}(t - t_j)\right]\right\}^{-1} & \text{for } t \geq t_j,\ j = 1, 2, \ldots, \end{cases} \quad (9)$$

where M_{Dj} and k_{Dj} denote the effective amplitude factor and elimination rate parameters, respectively, that now depend on the caffeine concentration at time t_j [26]. This repeated-dose g_{PD} model in Eq. (9) reduces to Eq. (7) under single-dose conditions.

The g_{PD} model in Eq. (9) does not consider the absorption of caffeine. This is a reasonable approximation for caffeine when ingested via coffee, tea, energy drinks, and most gum products, where the absorption rate is much faster (by a factor of >15) than the elimination rate.

4.1.2 Study Data

We obtained PVT lapse data from two previously published TSD studies to develop and validate the dose-dependent caffeine model.

The first study (labeled *Study C*) involved 48 healthy young adults who were kept awake for 29 consecutive hours in a controlled laboratory environment [27]. The 48 subjects were randomly assigned to one of four dose groups (placebo, 50, 100, or 200 mg, n = 12 subjects/group) and were administered the corresponding dose of Stay Alert® (Amurol Confectioners, Yorkville, IL) caffeinated chewing gum at the beginning of each of three 2-h test blocks after 20, 22, and 24 h of sleep loss (corresponding to 0300, 0500, and 0700 hours, respectively, on *day 2*). All subjects completed 10-min PVTs starting at 0800 hours on *day 1* and ending at 1200 hours on *day 2*, for a total of 29 PVT sessions, including nine sessions before caffeine administration, six sessions during each of the three subsequent 2-h test blocks, and two additional tests after the third 2-h test block.

The data from the second study (labeled *Study D*) were collected as part of a randomized Latin square crossover experiment across four laboratory sessions, each separated by at least 1 mo (washout period), in which 16 healthy young adults were kept awake for 27 consecutive hours [28]. During each of the four laboratory sessions, subjects were administered placebo, 100, 200, or 300 mg of Stay Alert® caffeinated chewing gum three times (the same dose of caffeine was administered in each of the three times) after 20, 22, and 24 h of sleep loss (corresponding to 0300, 0500, and 0700 hours, respectively, on *day 2*). Subjects completed 10-min PVTs starting at 0800 hours on *day 1* and ending at 1000 hours on *day 2*, for a total of 27 PVT sessions, including nine sessions before caffeine administration and six sessions after each of the three caffeine gum administrations.

For *Study C*, two subjects (one subject from the placebo group and one subject from the 100-mg group) were excluded from analyses due to missing data, resulting in a sample size of 11 subjects for the placebo and 100-mg groups. Three subjects from *Study D* (crossover design) were excluded due to missing data, resulting in a sample size of 13 subjects in this study.

Because baseline measures of performance (i.e., the first ~20 h involving the 9 sessions before caffeine administration) varied across *Studies C* and *D*, to allow for proper cross-study predictions, we normalized the performance data to eliminate baseline differences between the two studies. To this end, we applied an affine transformation [26, 29] to *Study D* data.

4.1.3 Group-average and Individual-specific Models

We used the group-average PVT lapse data from *Study C* to develop the group-average dose-dependent caffeine model. Specifically, we first obtained a group-average caffeine-free model $\bar{P}$ by fitting Eq. (1) to data from the

0-mg group of subjects. We then obtained the group-average caffeine model parameters (M_0, k_0, and z), needed to estimate the dose-dependent group-average $\bar{g}_{PD}$, by minimizing the combined sum of the squared errors between the caffeine model and the data from the different caffeine dose groups in the study. We then used this model to predict the group-average lapse data from *Study D*.

We used each individual's PVT lapse from *Study D* to develop individual-specific prediction models of caffeine effects. Specifically, for each subject *i*, we first obtained the caffeine-free component of the model P^i by fitting Eq. (1) to the *i*-th subject's performance data obtained under placebo (0 mg) administration. We then computed a group-average $\bar{g}^i_{PD}(t, D_j)$ using the same approach as described above. However, in computing the group-average $\bar{g}^i_{PD}$, we excluded performance data from the *i*-th subject. Accordingly, to predict performance of the *i*-th subject at time *t* after the *j*-th caffeine dose D_j, we used the following equation:

$$P^i_c(t, D_j) = P^i(t) \times \bar{g}^i_{PD}(t, D_j). \quad (10)$$

4.1.4 Goodness of Fits

To assess the goodness of fits, we calculated the RMSEs between the group-average caffeine model fits and predictions (and the individual-specific model predictions) and the performance data.

4.2 Results: Modeling Dose-dependent Effects of Caffeine

We used group-average lapse data from *Study C* to obtain the dose-dependent caffeine model parameters (M_0, k_0, and z) and then assessed the corresponding group-average model fits and cross-study predictions of *Study D* lapse data. Finally, we used *Study D* lapse data to construct individual-specific caffeine-free models P^i, individual-specific caffeine models P^i_c, and group-average caffeine models ($\bar{P}^i_c$ [$= \bar{P} \times \bar{g}^i_{PD}$]).

4.2.1 Group-average Model Fits and Predictions

We obtained the following dose-dependent caffeine model parameter estimates (95% CIs) by fitting Eq. (6) on *Study C* data: M_0 = 9.86 (6.96-13.98) lapses, k_0 = 0.49 (0.28-0.88) h^{-1}, and z = 1.63 (0.59-4.46) g^{-1}.

Figure 5, *top*, shows group-average PVT lapse data for *Study C* in the three caffeine dose groups (50, 100, and 200 mg; one in each panel) and the group-average model fits on these data. It also shows the group-average caffeine-free model $\bar{P}$, which was obtained by fitting $P(t)$ in Eq. (1) to the placebo data averaged across all subjects in the study. The caffeine model accurately captured the dose-dependent effects of the three repeated doses of caffeine on performance and showed considerably improved fits to the data compared with the caffeine-free model. Figure 5, *bottom*, shows group-average lapse data for *Study D* in the 100-, 200-, and 300-mg caffeine dose groups (one in each panel) and model predictions based on a group-average model developed on the entire *Study C* data. The caffeine model accurately predicted the effects of the three repeated doses of 100, 200, and 300 mg of caffeine and was substantially better than the caffeine-free model predictions.

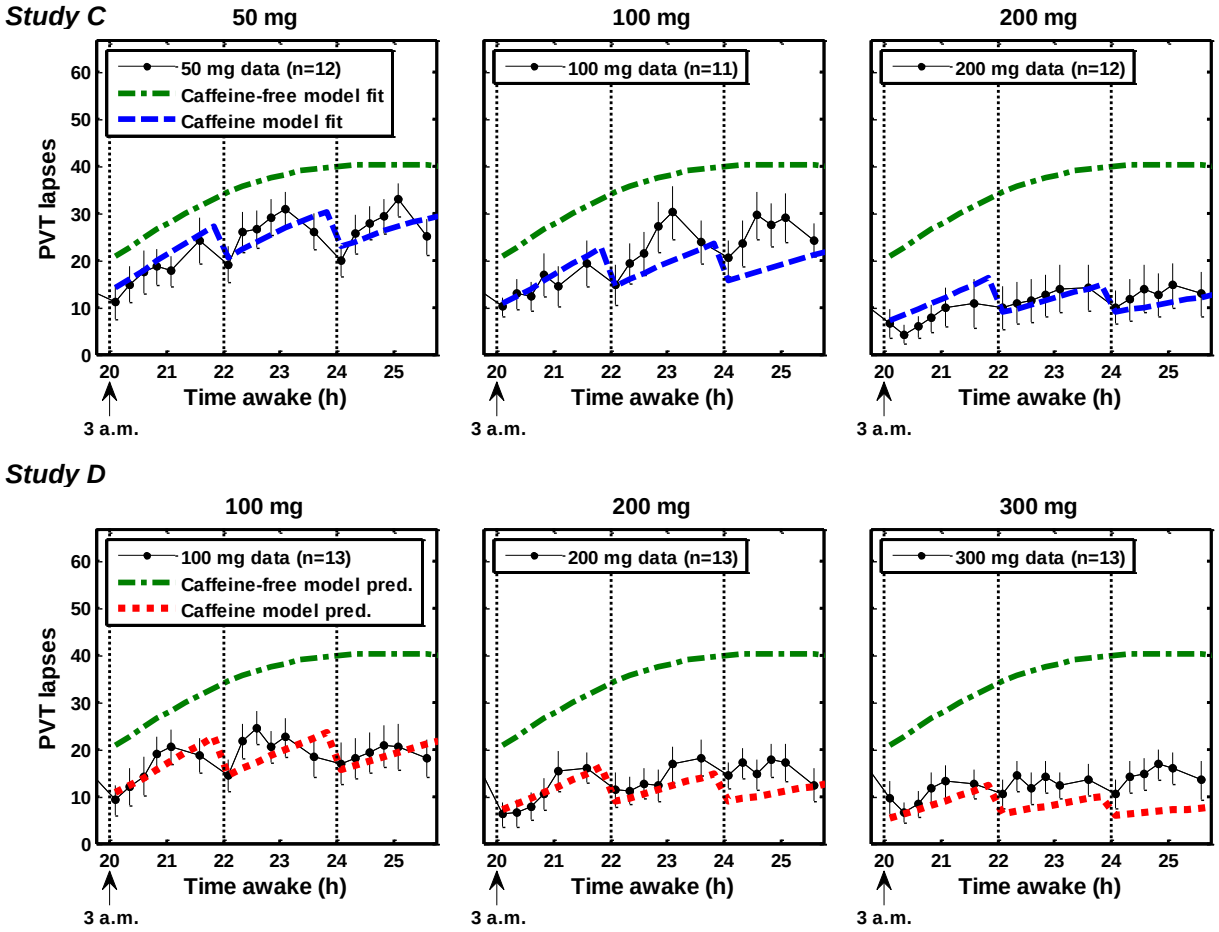

Figure 5: Dose-dependent caffeine model fits (*top*) and cross-study predictions (*bottom*) on group-average PVT lapse data (standard errors) measured after three repeated caffeine dose administrations (at 20, 22, and 24 h of sleep loss; denoted by thin dotted vertical lines) in *Studies C* (*top*) and *D* (*bottom*). The green dashed-dotted lines represent the corresponding caffeine-free model fits and predictions.

We assessed the group-average caffeine model fits (and predictions) by calculating the RMSEs between the fits (and predictions) and the group-average lapse data for each of the three caffeine dose groups in each study. Table 2 lists RMSEs of the group-average model fits on *Study C* and predictions on *Study D*. It also lists the associated RMSEs of the caffeine-free model within parentheses. For each dosing condition, the RMSEs of the caffeine model were at least 2.3 times smaller than their caffeine-free counterparts. In particular, the caffeine model showed 57-90% improvements over the caffeine-free model fits and 75-82% improvements over the caffeine-free model predictions.

Table 2: Root mean squared errors (RMSEs) of the caffeine model fits (for *Study C*) and predictions (for *Study D*) on group-average PVT lapse data. Entries within parentheses reflect the RMSEs of the corresponding caffeine-free models fits and predictions. RMSE units are number of PVT lapses.

Dose (mg)	Fit on *Study C*	Prediction on *Study D*
50	3.12 (11.46)	–
100	5.94 (13.83)	3.08 (16.86)
200	2.44 (24.01)	3.90 (21.58)
300	–	5.70 (22.51)

4.2.2 Individual-specific Model Predictions

Using the caffeine model in Eq. (10), we also developed individual-specific caffeine models P_c^i to predict post-caffeine performance of each subject *i* in *Study D* (crossover design study). We compared these predictions with the corresponding individual-specific, caffeine-free model estimates P^i [in Eq. (10) used to compute P_c^i] and the group-average caffeine model predictions $\overline{P}_c^i$ [$= \overline{P} \times \overline{g}_{PD}^i$ in Eq. (10) based on *Study D* data while excluding data from the *i*-th subject]. Figure 6 shows the P_c^i predictions after three repeated caffeine doses (the same dose of caffeine administered each of the three times) of 100, 200, and 300 mg for one of the 13 subjects (Subject #1) in *Study D*. Also shown are the individual-specific, caffeine-free model predictions P^i and group-average model predictions $\overline{P}_c^i$. Both P_c^i and $\overline{P}_c^i$ predicted the effects of 200 and 300 mg of caffeine considerably better than P^i. For the 100-mg caffeine doses, performance improved immediately after caffeine intake but dissipated quickly after ~30 min.

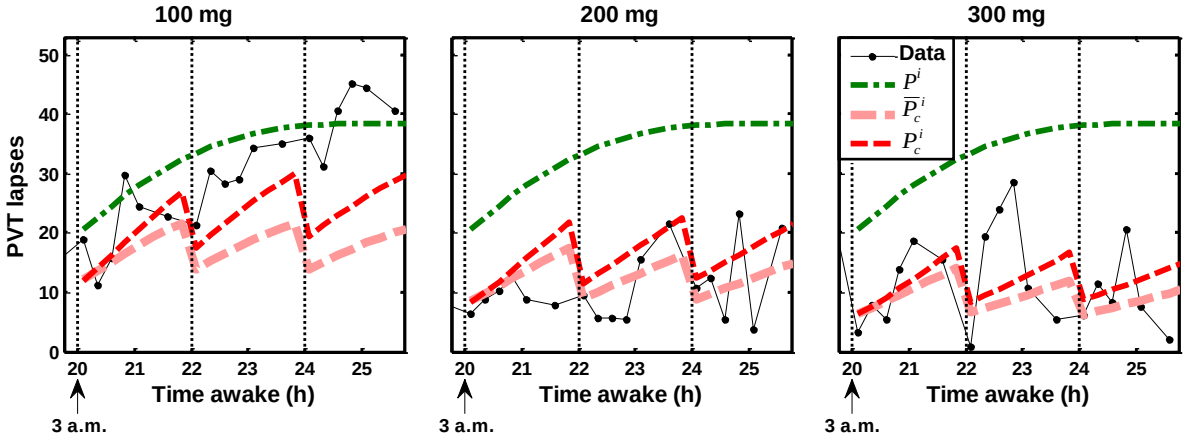

Figure 6: Individual-specific (P_c^i) and group-average ($\overline{P}_c^i$) caffeine model predictions of one of the subject's (Subject #1) PVT lapse data after three repeated caffeine doses of 100 mg (*left*), 200 mg (*middle*), and 300 mg (*right*) administered at 20, 22, and 24 h of sleep loss (denoted by thin dotted vertical lines) in *Study D* [28]. The green dashed-dotted lines represent the individual-specific caffeine-free model fit (P^i) on PVT lapse data obtained under placebo administration.

Table 3 shows a comparison of RMSEs of the individual-specific caffeine model predictions P_c^i with those of the corresponding caffeine-free model estimates P^i to assess the benefit of accounting for the effects of caffeine

doses of 100, 200, or 300 mg on performance in *Study D* subjects. The table also lists the median percent differences between the RMSEs of P_c^i and P^i across the 13 subjects for each dose. P_c^i performed better than P^i in at least 7 of 13 subjects in each dosing condition, with median improvements ranging from 31% for low (100 mg) caffeine doses to 43% for high (300 mg) doses.

4.3 Discussion: Modeling Dose-dependent Effects of Caffeine

The mathematical model described here captures the dose-dependent effects of caffeine on PVT lapses during total sleep loss. It builds on our previously developed fixed-dose model [18], which estimates the effect of caffeine by multiplying the phenomenological two-process model of sleep regulation (used to characterize performance in the absence of caffeine) with a caffeine-effect factor (g_{PD}) that ranges from 0 (maximal caffeine effect) to 1 (no caffeine effect). The g_{PD} factor is based on the PK-PD sigmoidal relationship of caffeine derived via the Hill equation. Here, we incorporated dose-dependent effects into g_{PD} by modeling two of its parameters, which control the strength and duration of caffeine effects, as a function of dose. Specifically, we modeled g_{PD} such that the strength of caffeine effect increases with larger doses, but only up to a point (best PVT performance), beyond which the magnitude of the effect saturates. However, the duration of effect keeps increasing with dose, reflecting the saturable metabolic processes involved in the clearance of caffeine [22, 30].

Table 3: RMSEs of the individual-specific caffeine model predictions (P_c^i) and individual-specific caffeine-free model estimates (P^i) of PVT lapse data for *Study D* subjects after three repeated caffeine doses of 100, 200, and 300 mg administered at 20, 22, and 24 h of sleep loss. Also shown for each dose are the median percent differences [= 100 x ($P^i - P_c^i$)/ P^i] between the RMSEs across all 13 subjects. RMSEs of caffeine models that performed better than their corresponding caffeine-free models are in boldface. RMSE units are number of PVT lapses.

Subject	*100 mg*		*200 mg*		*300 mg*	
	P_c^i	P^i	P_c^i	P^i	P_c^i	P^i
1	10.85	6.31	**6.72**	23.42	**7.69**	23.35
2	**5.30**	7.65	**6.40**	9.94	**8.41**	8.45
3	11.68	8.31	**11.28**	20.83	16.13	14.69
4	19.61	12.38	**9.63**	9.80	**9.09**	15.89
5	**16.96**	25.40	**14.24**	29.46	**10.56**	29.82
6	**9.18**	14.62	13.57	13.00	21.94	8.29
7	19.63	6.99	20.19	6.22	**11.81**	20.24
8	**10.50**	15.19	15.30	9.36	**7.28**	10.34
9	**14.84**	21.94	**9.39**	21.21	**8.28**	15.46
10	**11.14**	19.05	**11.91**	23.84	**6.41**	15.51
11	**11.65**	17.65	**9.34**	19.85	**6.99**	20.26
12	10.83	4.02	9.34	9.29	16.12	4.60
13	7.62	7.48	**9.79**	10.41	**5.46**	16.26
Median % Difference	31		36		43	

A group-average, dose-dependent caffeine model was developed on data from one study and then validated on data from another study. The caffeine model captured (and predicted) the dose-dependent effects of caffeine on performance in both studies. Importantly, the model fits and cross-study predictions were substantially better (up to 90% and 82%, respectively) than the caffeine-free models (Table 2), with greater improvements typically observed in larger doses.

Individual-specific caffeine models were also developed, where the caffeine-free component was individualized and the caffeine-effect multiplier g_{PD} was based on the group-average model. The individual-specific caffeine models yielded an average reduction in prediction error of 37% compared with their caffeine-free counterparts across all doses, with improvements being greater for larger doses.

An expanded treatment of these findings was previously reported in Ref. 26.

5.0 CONCLUSIONS

In summary, in this work, we have developed key capabilities to accurately predict the effects of sleep loss and the restorative effects of different dosages of caffeine on performance at both group-average and individual levels. To enhance the utility of the models, we seek to incorporate additional capabilities. In particular, we are developing strategies to integrate the UMP with the dose-dependent model of the caffeine response so as to predict the detrimental effects of both total and partial sleep loss and the recuperative effects of caffeine using a single model. This supports our long-term goal of incorporating these model components into an integrated computational tool that prescribes countermeasures (e.g., the timing of naps and timing and dosage of caffeine) to optimize an individual's neurobehavioral performance and thereby reduce the risk of sleep-loss-related errors and accidents.

While many challenges remain, the integrated UMP would provide another step toward the development of a wearable computer-based system or smartphone app that considers an individual's sleep/wake history, current and recent-past performance, and caffeine consumption to predict future levels of performance.

6.0 ACKNOWLEDGEMENTS

This work was sponsored by the Military Operational Medicine Research Area Directorate of the U.S. Army Medical Research and Materiel Command, Ft. Detrick, MD, and by the U.S. Department of Defense Medical Research and Development Program (Grant No. DMRDP_13200).

7.0 DISCLAIMER

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution.

8.0 REFERENCES

[1] U. S. Army, *Report of the Mental Health Advisory Team 9 (MHAT 9), Operation Enduring Freedom (OEF) 2013, Afghanistan*: DIANE Publishing Company, 2014.

[2] M. M. Mallis, S. Mejdal, T. T. Nguyen, and D. F. Dinges, "Summary of the key features of seven biomathematical models of human fatigue and performance," *Aviat Space Environ Med*, vol. 75, pp. A4-14, Mar 2004.

[3] A. A. Borbely, "A two process model of sleep regulation," *Hum Neurobiol*, vol. 1, pp. 195-204, 1982.

[4] M. L. Johnson, G. Belenky, D. P. Redmond, D. R. Thorne, J. D. Williams, S. R. Hursh, and T. J. Balkin, "Modulating the homeostatic process to predict performance during chronic sleep restriction," *Aviat Space Environ Med*, vol. 75, pp. A141-6, Mar 2004.

[5] P. Achermann and A. A. Borbely, "Combining different models of sleep regulation," *J Sleep Res*, vol. 1, pp. 144-147, Jun 1992.

[6] P. Rajdev, D. Thorsley, S. Rajaraman, T. L. Rupp, N. J. Wesensten, T. J. Balkin, and J. Reifman, "A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation," *J Theor Biol*, vol. 331, pp. 66-77, Apr 24 2013.

[7] D. F. Dinges and J. W. Powell, "Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations," *Behav Res Methods Instrum Comput*, vol. 17, pp. 652-55, Dec 1985.

[8] J. Dorrian, N. L. Rogers, and D. F. Dinges, "Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss," in *Sleep deprivation. Clinical issues, pharmacology, and sleep loss effects.*, C. A. Kushida (Ed.), pp. 39-70. New York: Marcel Dekker, 2005.

[9] G. Belenky, N. J. Wesensten, D. R. Thorne, M. L. Thomas, H. C. Sing, D. P. Redmond, M. B. Russo, and T. J. Balkin, "Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study," *J Sleep Res*, vol. 12, pp. 1-12, Mar 2003.

[10] T. L. Rupp, N. J. Wesensten, and T. J. Balkin, "Trait-like vulnerability to total and partial sleep loss," *Sleep*, vol. 35, pp. 1163-72, Aug 2012.

[11] C. D. Schunn and D. Wallach, "Evaluating goodness-of-fit in comparison of models to data," *Psychologie der Kognition: Reden and vorträge anlässlich der emeritierung von Werner Tack*, pp. 115-54, 2005.

[12] H. P. Van Dongen, M. D. Baynard, G. Maislin, and D. F. Dinges, "Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability," *Sleep*, vol. 27, pp. 423-33, May 1 2004.

[13] T. Rusterholz, R. Durr, and P. Achermann, "Inter-individual differences in the dynamics of sleep homeostasis," *Sleep*, vol. 33, pp. 491-8, Apr 1 2010.

[14] J. H. Zar, *Biostatistical Analysis*. Upper Saddle River: Prentice-Hall, Inc., 1999.

[15] S. Ramakrishnan, W. Lu, S. Laxminarayan, N. J. Wesensten, T. L. Rupp, T. J. Balkin, and J. Reifman, "Can a mathematical model predict an individual's trait-like response to both total and partial sleep loss?," *J Sleep Res*, vol. 24, pp. 262-9, Jun 2015.

[16] M. Basner, D. Mollicone, and D. F. Dinges, "Validity and Sensitivity of a Brief Psychomotor Vigilance Test (PVT-B) to Total and Partial Sleep Deprivation," *Acta Astronaut*, vol. 69, pp. 949-59, Dec 1 2011.

[17] M. Y. Khitrov, S. Laxminarayan, D. Thorsley, S. Ramakrishnan, S. Rajaraman, N. J. Wesensten, and J. Reifman, "PC-PVT: a platform for psychomotor vigilance task testing, analysis, and prediction," *Behav Res Methods*, vol. 46, pp. 140-7, Mar 2014.

[18] S. Ramakrishnan, S. Rajaraman, S. Laxminarayan, N. J. Wesensten, G. H. Kamimori, T. J. Balkin, and J. Reifman, "A biomathematical model of the restoring effects of caffeine on cognitive performance during sleep deprivation," *J Theor Biol*, vol. 319, pp. 23-33, Feb 21 2013.

[19] M. Bonati, R. Latini, F. Galletti, J. F. Young, G. Tognoni, and S. Garattini, "Caffeine disposition after oral doses," *Clin Pharmacol Ther*, vol. 32, pp. 98-106, Jul 1982.

[20] G. H. Kamimori, C. S. Karyekar, R. Otterstetter, D. S. Cox, T. J. Balkin, G. L. Belenky, and N. D. Eddington, "The rate of absorption and relative bioavailability of caffeine administered in chewing gum versus capsules to normal healthy volunteers," *Int J Pharm*, vol. 234, pp. 159-67, Mar 2 2002.

[21] C. Csajka and D. Verotta, "Pharmacokinetic-pharmacodynamic modelling: history and perspectives," *J Pharmacokinet Pharmacodyn*, vol. 33, pp. 227-79, Jun 2006.

[22] C. P. Denaro, C. R. Brown, M. Wilson, P. Jacob, 3rd, and N. L. Benowitz, "Dose-dependency of caffeine metabolism with repeated dosing," *Clin Pharmacol Ther*, vol. 48, pp. 277-85, Sep 1990.

[23] G. H. Kamimori, S. I. Lugo, D. M. Penetar, A. C. Chamberlain, G. E. Brunhart, A. E. Brunhart, and N. D. Eddington, "Dose-dependent caffeine pharmacokinetics during severe sleep deprivation in humans," *Int J Clin Pharmacol Ther*, vol. 33, pp. 182-6, Mar 1995.

[24] G. B. Kaplan, D. J. Greenblatt, B. L. Ehrenberg, J. E. Goddard, M. M. Cotreau, J. S. Harmatz, and R. I. Shader, "Dose-dependent pharmacokinetics and psychomotor effects of caffeine in humans," *J Clin Pharmacol*, vol. 37, pp. 693-703, Aug 1997.

[25] M. Gibaldi and D. Perrier, *Pharmacokinetics*. New York: Marcel Dekker, 1982.

[26] S. Ramakrishnan, S. Laxminarayan, N. J. Wesensten, G. H. Kamimori, T. J. Balkin, and J. Reifman, "Dose-dependent model of caffeine effects on human vigilance during total sleep deprivation," *J Theor Biol*, vol. 358, pp. 11-24, 2014.

[27] G. H. Kamimori, D. Johnson, D. Thorne, and G. Belenky, "Multiple caffeine doses maintain vigilance during early morning operations," *Aviat Space Environ Med*, vol. 76, pp. 1046-50, Nov 2005.

[28] C. M. LaJambe, G. H. Kamimori, G. Belenky, and T. J. Balkin, "Caffeine effects on recovery sleep following 27 h total sleep deprivation," *Aviat Space Environ Med*, vol. 76, pp. 108-13, Feb 2005.

[29] T. J. Hastie, R. J. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer, 2001.

[30] W. S. Cheng, T. L. Murphy, M. T. Smith, W. G. Cooksley, J. W. Halliday, and L. W. Powell, "Dose-dependent pharmacokinetics of caffeine in humans: relevance as a test of quantitative liver function," *Clin Pharmacol Ther*, vol. 47, pp. 516-24, Apr 1990.