Journal of Theoretical Biology 331 (2013) 66–77

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation

Pooja Rajdev [a], David Thorsley [a], Srinivasan Rajaraman [a], Tracy L. Rupp [b], Nancy J. Wesensten [b], Thomas J. Balkin [b], Jaques Reifman [a,*]

[a] Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, United States Army Medical Research and Materiel Command, Fort Detrick, MD, USA
[b] Department of Behavioral Biology, Walter Reed Army Institute of Research, Silver Spring, MD, USA

HIGHLIGHTS

- We developed a model of neurocognitive performance that incorporates sleep debt.
- Our model unifies total sleep deprivation and chronic sleep restriction scenarios.
- Our model captures the slower recovery process after chronic sleep restriction.
- Our model describes the beneficial effects of banking sleep on performance.

ARTICLE INFO

Article history:
Received 19 October 2012
Received in revised form
28 February 2013
Accepted 12 April 2013
Available online 24 April 2013

Keywords:
Biomathematical models
Two-process model of sleep regulation
Neurocognitive performance
Chronic sleep restriction
Total sleep deprivation

ABSTRACT

Performance prediction models based on the classical two-process model of sleep regulation are reasonably effective at predicting alertness and neurocognitive performance during total sleep deprivation (TSD). However, during sleep restriction (partial sleep loss) performance predictions based on such models have been found to be less accurate. Because most modern operational environments are predominantly characterized by chronic sleep restriction (CSR) rather than by episodic TSD, the practical utility of this class of models has been limited.

To better quantify performance during both CSR and TSD, we developed a unified mathematical model that incorporates extant sleep debt as a function of a known sleep/wake history, with recent history exerting greater influence. This incorporation of sleep/wake history into the classical two-process model captures an individual's capacity to recover during sleep as a function of sleep debt and naturally bridges the continuum from CSR to TSD by reducing to the classical two-process model in the case of TSD. We validated the proposed unified model using psychomotor vigilance task data from three prior studies involving TSD, CSR, and sleep extension. We compared and contrasted the fits, within-study predictions, and across-study predictions from the unified model against predictions generated by two previously published models, and found that the unified model more accurately represented multiple experimental studies and consistently predicted sleep restriction scenarios better than the existing models. In addition, we found that the model parameters obtained by fitting TSD data could be used to predict performance in other sleep restriction scenarios for the same study populations, and vice versa. Furthermore, this model better accounted for the relatively slow recovery process that is known to characterize CSR, as well as the enhanced performance that has been shown to result from sleep banking.

Published by Elsevier Ltd.

1. Introduction

Sleepiness increases the risk of human error and accidents. It also affects the health, safety, and quality of life of military and civilian personnel who are regularly exposed to work schedules that preclude adequate daily sleep duration and timing (Mallis et al., 2004). Critical to effective management of operational alertness and performance is the ability to accurately predict the impact of various work/rest schedules on individual operators. In

Abbreviations: CSR, chronic sleep restriction; EV, explained variance; PVT, psychomotor vigilance test; R^2, adjusted coefficient of determination; RMSE, root mean squared error; RRT, reciprocal response time; RT, response time; RTD, response time divergence; SWT, scheduled wake time; TIB, time in bed; TSD, total sleep deprivation

* Correspondence to: Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, United States Army Medical Research and Materiel Command, ATTN: MCMR-TT, 504 Scott Street, Fort Detrick, MD 21702, USA.
Tel.: +1 301 619 7915; fax: +1 301 619 1983.
E-mail address: jaques.reifman.civ@mail.mil (J. Reifman).

0022-5193/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.jtbi.2013.04.013

this paper, we consider the problem of predicting the alertness and performance of a population for a set sleep/wake schedule. Biomathematical modeling provides the most promising strategy for addressing the problem of helping manage alertness and neurocognitive performance in operational environments (Friedl et al., 2004), thereby enhancing the safety and productivity of both military and civilian operators.

Borbély's seminal two-process model, originally developed to describe the mechanisms mediating sleep regulation (Borbély, 1982), has also served as the basis of many models used to predict human alertness and neurocognitive performance during sleep loss (Mallis et al., 2004). A basic postulate of this model is that alertness and performance are modulated by the additive interaction of two processes. The first, process *S*, is the sleep homeostat, responsible for increasing sleep propensity during waking, and reducing sleep propensity as recovery occurs during sleep. The fluctuations of *S* are described by exponential functions with fixed upper and lower asymptotes. The second process is the endogenous circadian rhythm, process *C*, which is driven by the internal clock residing in the suprachiasmatic nuclei of the anterior hypothalamus (Daan et al., 1984). This phenomenological model, based on findings from acute total sleep deprivation (TSD) studies, has been extended beyond its original goal of predicting slow-wave activity as a function of sleep/wake history, and now also provides a theoretical framework for quantifying the effects of sleep deprivation on objective and subjective alertness and neurocognitive performance.

More recently, several groups have investigated performance degradation resulting from different levels of sleep restriction (Belenky et al., 2003; Van Dongen et al., 2003; Rupp et al., 2009, 2012; Carskadon and Dement, 1981; Dinges et al., 1997). In contrast to TSD studies, results from these well-controlled chronic sleep restriction (CSR) studies have shown that models of neurocognitive performance based solely on Borbély's two-process model fail to accurately predict the observed degradation (Van Dongen et al., 2003; Carskadon and Dement, 1981; Dinges et al., 1997; Mollicone et al., 2010). It has also been observed that the rate of neurocognitive performance recovery after CSR is considerably slower than the rate of recovery after acute TSD (Belenky et al., 2003; Johnson et al., 2004) and that this class of models does not accurately capture this difference (Johnson et al., 2004).

New performance prediction models have been proposed to explain these observations. Van Dongen et al. (2003) described the effects of sleep restriction in terms of "excess wakefulness" or cumulative wake-time extensions rather than as a homeostatic process, whereas Johnson et al. (2004) and Hursh et al. (2004) introduced a "slow" process modulating the homeostat based on sleep/wake history. The latter model provides accurate predictions for aggregate, daily mean performance during seven days of CSR. However, it has not been used to describe performance variations within each day. Avinash et al. (2005) used this slow process to manipulate the upper and lower asymptotes of the sleep homeostat process in the two-process model so that they simultaneously rise during wakefulness and decay during sleep while maintaining a constant, fixed difference between them. A limitation of this approach is the requirement for an *a priori* estimate of the exact value of the fixed difference between the asymptotes, which is likely to vary across different data sets. In addition, they found that although these models accurately fit data collected under CSR, they substantially underestimate performance impairment under TSD Avinash et al. (2005).

McCauley et al. (2009) showed that these approaches belong to a broader class of homeostatic models and incorporated the two-process model and Avinash et al.'s model into a generalized state-space model. This state-space model similarly maintains a constant, fixed difference between the homeostat asymptotes as they rise and fall. In addition, it predicts a bifurcation of the performance trajectory; that is, when daily wakefulness is maintained below a critical threshold, performance tends to stabilize at a deteriorated level, whereas when daily wakefulness is increased beyond this threshold, the model predicts a continuous degradation in performance over time. This predicted bifurcation of the performance trajectory follows a timescale much longer than the duration of their 14-day CSR study on which the model was based and thus has not been experimentally validated. Moreover, the inclusion of seven additional parameters beyond the seven parameters of the classical two-process model (Avinash et al., 2005; Achermann and Borbély, 1994) makes it difficult to estimate the model parameters from limited CSR data. Furthermore, the inherently nonlinear interaction between the homeostatic and circadian processes in this model can place the lower asymptote above actual performance data.

To address these limitations, we developed a model we call the "unified model." Results of recent studies suggest that CSR induces relatively long-term, slow-recovering changes in brain physiology that affect alertness and performance (Belenky et al., 2003; Johnson et al., 2004; Alhola and Polo-Kantola, 2007). We hypothesized that these long-term changes alter the homeostatic process during sleep such that the capacity of an individual to recover during sleep changes as a function of prior sleep/wake history, i.e., as a function of sleep debt. Mathematically, we modeled this hypothesis by allowing the lower asymptote of the classical two-process model to increase or decrease based on the accumulation or restoration of sleep debt, respectively, while keeping the upper asymptote constant. Because the lower asymptote bounds performance impairment from below, constraining the minimum amount of impairment, by modulating the lower asymptote as a function of sleep debt we effectively constrain the rate of performance recovery during sleep. Sleep debt, in turn, is modeled based on a "fading memory" filter, representing the notion that sleep losses or sleep extensions that occurred in the remote past have less effect on the present sleep debt and performance than comparable events in the recent past. Belenky et al. (2003) proposed a similar notion to explain the slow rate of performance recovery after CSR; our work builds on their observations by constructing a mathematical model to describe the phenomenon. A similar notion of fading memory is used in the Fatigue Audit InterDyne (FAID) model developed by Dawson and Fletcher (2001); the fading-memory filter we propose in this paper goes beyond the FAID model in that it also incorporates the possible beneficial effects of sleep banking (Rupp et al., 2009). The idea of fading memory has not been incorporated into Borbély's two-process model (and other closely related neurocognitive performance models), which assume a constant capacity to recover from sleep loss regardless of prior sleep/wake history.

The unified model is so named because it bridges the continuum between CSR and TSD and reduces to Borbély's classical two-process model in the case of total sleep loss. We validated the proposed model using data from three prior studies (Belenky et al., 2003; Van Dongen et al., 2003; Rupp et al., 2012) in which subjects were exposed to TSD as well as different CSR schedules.

2. Methods

2.1. Borbély's two-process model of sleep regulation

Borbély's two-process model (Borbély and Achermann, 1999; Achermann and Borbély, 1992) is based on the interaction of two processes: (1) the homeostatic process *S*, which rises monotonically during wakefulness and declines monotonically during sleep (Daan et al., 1984) and (2) a circadian process *C*, which is a 24-h

periodic, self-sustaining oscillator modeled as a five-harmonic sinusoidal equation.

The homeostatic process $S(t)$ is defined for each time t by two exponential functions, one each for wakefulness and sleep (Borbély and Achermann, 1999; Achermann and Borbély, 1992):

$$\dot{S}(t) = [U - S(t)]/\tau_w \quad \text{during wakefulness} \tag{1a}$$

$$\dot{S}(t) = -[S(t) - L]/\tau_s \quad \text{during sleep} \tag{1b}$$

where U represents the upper asymptote, L represents the lower asymptote, τ_w denotes the time constant of the increasing saturating exponential function during wakefulness, and τ_s denotes the time constant of the decreasing exponential function during sleep. In the classical two-process model, both U and L are constant for all t.

The circadian process $C(t)$ at time t is modeled by a five-harmonic sinusoidal equation (Daan et al., 1984):

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

where a_i, $i = 1,\ldots, 5$, represent the amplitude of the five harmonics of the circadian process ($a_1 = 0.97$, $a_2 = 0.22$, $a_3 = 0.07$, $a_4 = 0.03$, and $a_5 = 0.001$), τ denotes the period of the circadian oscillator (~24 h), and φ denotes the circadian phase (Achermann and Borbély, 1992). This five-harmonic sinusoidal equation models the circadian pacemaker under entrained conditions (Klerman and St-Hilaire, 2007).

We modeled the dynamics of neurocognitive performance as the additive interaction of the homeostatic process S and the circadian process C. Thus, performance impairment $P(t)$ at time t is given by

$$P(t) = S(t) + \kappa \times C(t) \tag{3}$$

where κ is a positive constant representing the relative effect of the circadian process on performance. Consequently, the seven parameters required for estimating performance during intermittent periods of sleep and wakefulness using the two-process model are $S(0)$, $U(0)$, $L(0)$, τ_w, τ_s, φ, and κ. In this formulation, large values of $P(t)$, $S(t)$, and $C(t)$ indicate greater performance impairment.

2.2. The state-space model

Avinash et al. (2005) extended the two-process model by allowing the asymptotes $L(t)$ and $U(t)$ to increase linearly during wakefulness and decrease exponentially during sleep while maintaining a constant difference between them at all times. McCauley et al. (2009) showed that this model belongs to a broader class of homeostatic models that can be represented by a system of coupled nonhomogeneous first-order ordinary differential equations. In this class of models, the asymptotes increase exponentially during wakefulness and decrease exponentially during sleep, and this general model is expressed as

$$\begin{bmatrix} \dot{P}_w \\ \dot{U} \end{bmatrix} = \underline{\mathbf{A}} \begin{bmatrix} P_w \\ U \end{bmatrix} + \begin{bmatrix} \mu + \kappa \times C(t) \\ 0 \end{bmatrix},$$

$$\underline{\mathbf{A}} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \quad \text{during wakefulness} \tag{4a}$$

$$\begin{bmatrix} \dot{P}_s \\ \dot{L} \end{bmatrix} = \underline{\mathbf{\Sigma}} \begin{bmatrix} P_s \\ L \end{bmatrix} + \begin{bmatrix} \mu + \kappa \times C(t) \\ 0 \end{bmatrix},$$

$$\underline{\mathbf{\Sigma}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \quad \text{during sleep} \tag{4b}$$

$$L(t) = U(t) - \delta, \quad \text{during wakefulness and sleep} \tag{4c}$$

where P_w and P_s denote performance values during wakefulness and sleep, respectively, $\underline{\mathbf{A}}$ represents a 2×2 matrix that modulates P_w and the upper asymptote U during wakefulness, $\underline{\mathbf{\Sigma}}$ represents a 2×2 matrix that modulates P_s and the lower asymptote L during sleep, μ and κ represent scaling parameters for the circadian process, and δ denotes the constant difference between the asymptotes. The diagonal elements of matrix $\underline{\mathbf{A}}$, α_{11} and α_{22}, represent the time constants of processes P_w and U, respectively, and the off-diagonal elements, α_{12} and α_{21}, represent the influences of U on P_w and of P_w on U, respectively. Similar relations exist between the elements of matrix $\underline{\mathbf{\Sigma}}$ and P_s and L during sleep. This state-space model contains a total of 14 parameters. However, in the particular case modeled in McCauley et al. (2009), it is assumed that performance does not affect the modulation of the asymptotes. Therefore, α_{21} and σ_{21} are set to 0 and the number of model parameters is reduced to 12.

The homogenous part of the differential equation (i.e., the first term on the right-hand side of Eqs. (4a) and (4b)) defines the homeostatic process and governs the dynamics of the model over a scale of days, whereas the nonhomogeneities (i.e., the second term on the right-hand side of Eqs. (4a) and (4b)) define the circadian process and govern the changes within the course of a day. In this representation, the two model components interact in a nonlinear fashion. This class of models predicts the existence of a bifurcation point, or a critical threshold at $T\sigma_{22}/(\sigma_{22} - \alpha_{22})$ hours of daily wakefulness, where T denotes the 24-h wake/sleep cycle duration. When the amount of daily wakefulness is less than this critical threshold, performance is predicted to stabilize, whereas when the duration of daily wakefulness is greater than the critical threshold, performance is predicted to continuously degrade without limit.

2.3. The unified model

Our proposed model extends Borbély's two-process model by taking into account a known sleep/wake history and sleep debt in order to better describe and predict performance during chronic sleep restriction and subsequent recovery sleep. The sleep debt of an individual on any day is defined using a recursive filter that incorporates an exponential decay of the sleep-loss history. Specifically, the amount of sleep loss $Loss(t)$ and the rate of change of sleep debt $dDebt(t)/dt$ at an instant t are defined as

$$Loss(t) = \begin{cases} 1 & \text{while awake} \\ -2 & \text{while asleep} \end{cases} \tag{5a}$$

$$\frac{dDebt(t)}{dt} = \frac{-Debt(t) + Loss(t)}{\tau_{LA}} \tag{5b}$$

where τ_{LA} denotes the time constant of the filter. The values of $Loss(t)$ during sleep and wake were chosen so that if an individual sleeps 8 h, the recommended optimum sleep time per night (Belenky et al., 2003), the area under the curve (AUC) for $Loss(t)$ over a 24-h period is equal to zero ($16 \times 1 + (-2) \times 8$). In the differential equation defining the dynamics of $Debt(t)$, sleep losses or sleep extensions that occurred in the remote past have a much weaker influence on the present sleep debt than comparable events in the more recent past. To initialize the model, we set $Debt(0)$ to any value between −2 and 1, which ensures that $Debt(t)$ lies in this range for all $t > 0$.

The unified model incorporates sleep debt into the classical two-process model by describing changes in an individual's capacity to recover during sleep as a function of $Debt(t)$, i.e., the lower asymptote $L(t)$ of the homeostatic process is allowed to increase or decrease with increased or decreased $Debt(t)$, respectively, while the upper asymptote U remains constant. Thus, the

dynamics of process S during sleep are expressed as

$$\frac{dS(t)}{dt} = \frac{-[S(t)-L(t)]}{\tau_S} \tag{5c}$$

where $L(t)=U\times Debt(t)$ changes continuously with time. The upper asymptote U is constant for all time t within a sleep deprivation or sleep restriction scenario. During wakefulness, $S(t)$ follows Eq. (1a), and thus the unified model simulates Borbély's two-process model in the case of TSD. During sleep, the differential equation formulation in Eqs. (5b) and (5c) can be solved explicitly, giving the solution

$$L(t) = \left\{ L(t_0)\exp\left(\frac{-(t-t_0)}{\tau_{LA}}\right) - 2U\left[1-\exp\left(\frac{-(t-t_0)}{\tau_{LA}}\right)\right]\right\} \tag{6a}$$

$$S(t) = S(t_0)\exp\left(\frac{-(t-t_0)}{\tau_s}\right) - 2U\left[1-\exp\left(\frac{-(t-t_0)}{\tau_s}\right)\right] + \frac{(L(t_0)+2U)\tau_{LA}}{\tau_{LA}-\tau_s}\left[\exp\left(\frac{-(t-t_0)}{\tau_{LA}}\right) - \exp\left(\frac{-(t-t_0)}{\tau_s}\right)\right] \tag{6b}$$

where t_0 denotes the time of sleep onset. The circadian process $C(t)$ and performance $P(t)$ in this model are given by Eqs. (2) and (3), respectively. This model only introduces one additional parameter, τ_{LA}, over the two-process model and, therefore, contains a total of eight parameters: $S(0)$, U, $L(0)$, τ_w, τ_s, φ, κ, and τ_{LA}.

For an individual restricted to 8-h TIB each day, the AUC for $Loss(t)$ each day would be zero, and the lower asymptote $L(t)$ in Eq. (5c) would oscillate around zero, reaching its highest point before bedtime and its lowest point at awakening. If the time constant τ_{LA} is slow (e.g., $\tau_{LA}=120$ h), the magnitude of this oscillation would be 0.06 U, and thus the unified model would closely approximate Borbély's two-process model. Fig. 1 shows simulated results that illustrate the dynamics of the homeostat S of the unified model for CSR (3-h TIB; Fig. 1A) and TSD (64 h; Fig. 1B), followed by three recovery nights (8-h TIB). Restricting an individual to 3-h TIB resulted in an AUC for $Loss(t)$ of 0.625 for each day (Eq. (5a)). Thus, $Debt(t)$ asymptotically approached a limit cycle around 0.625 (Eq. (5b)), and, after 7 days, the lower asymptote $L(t)$ reached 55.7% of the upper asymptote U (Fig. 1A) and the homeostat S (Eq. (5c)) increased to its highest value. Subsequently, when the individual was allowed 8-h TIB in three successive nights, $L(t)$ decreased rapidly and, given enough nights of 8-h TIB, it would eventually oscillate around zero. For the TSD scenario (Fig. 1B), the homeostat $S(t)$ rose exponentially and was not affected by the increase in the lower asymptote, as is the case in the classical two-process model. Then, when the individual was allowed 8-h TIB in successive recovery nights, $L(t)$ again decreased and asymptotically approached a limit cycle around zero.

After 7 days of 3-h TIB, the homeostat S reached its maximum value, $S_{max,A}$. This value was slightly less than the value $S_{max,B}$ reached by S after 64 h of TSD. However, after the first night of recovery sleep following CSR, the homeostat value $S_{rec,A}$ was greater than the homeostat value $S_{rec,B}$ after the first night of recovery sleep following TSD (Fig. 1A and B). This difference in the value of the homeostat after recovery occurred because the value of $L(t)$ after 7 days of 3-h TIB was approximately 55.7% of U, whereas after 64 h of TSD it was only approximately 45.2% of U. This indicates that $Debt(t)$ after 7 days of 3-h TIB was approximately 0.557, whereas that after 64 h of TSD was only approximately 0.452. This difference in sleep debt and, thus, the lower asymptote values, predicted that the immediate recovery in performance after 7 days of 3-h TIB would be slower than the recovery after 64 h of TSD. That is, a larger sleep debt is modeled as a higher value of $L(t)$, which constrains the amount of recovery during sleep.

2.4. Fitting and prediction

We used population-average performance data to estimate the model parameters by performing a nonlinear least-squares optimization to minimize the root mean squared error (RMSE) between the population-average data and the model outputs. We also used each model with parameters estimated from one study scenario to predict the population-average performance data in a different study scenario. This allowed us to investigate whether models based on one of the scenarios could simultaneously capture the effects of TSD, CSR, and variation in sleep/wake history on performance. Following (Rusterholz et al., 2010), we bounded τ_w to lie between 0 and 40 h and τ_s to be greater than 1 h. All calculations were performed in MATLAB R2012a.

2.5. Statistics

We used RMSE and the adjusted coefficient of determination (R^2) to evaluate the goodness of fit of each model. For a data set with N population-average performance observations and a model with p parameters, the R^2 of the fit is given as follows:

$$R^2 = 1-(1-r^2)\frac{N-1}{N-p-1}, \quad \text{with} \quad r^2 = 1-\frac{SS_{err}}{SS_{tot}}, \tag{6c}$$

where r^2 denotes the unadjusted coefficient of determination, SS_{err} represents the sum of squares of the error between the observations and the model fits over the N observations, and SS_{tot} represents the total sum of squares between each of the N observations and the mean value of all of the observations

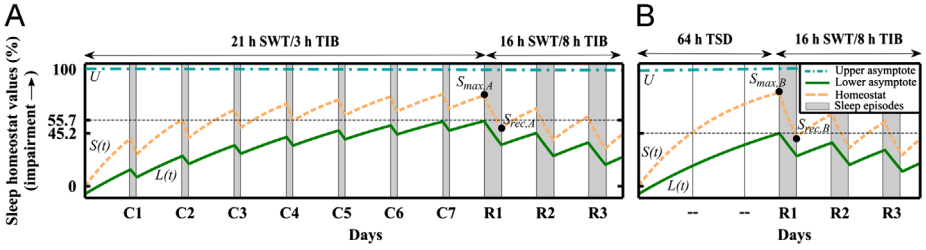

Fig. 1. Simulated dynamics of the sleep homeostatic process of the unified model under two different sleep/wake scenarios. A: When an individual was restricted to 21 h of scheduled wake time (SWT)/3 h of time in bed (TIB) for an extended period of time (i.e., seven nights, C1–C7), the lower asymptote ($L(t)$, solid green line) reached 55.7% of the upper asymptote (U, dashed-dotted blue line). This increase in the lower asymptote causes an effective decrease in the capacity of the individual to recover during sleep. When the individual was allowed 8-h TIB (three nights, R1–R3), the lower asymptote monotonically decreased, allowing the individual to recover. B: When an individual was exposed to total sleep deprivation (TSD), the homeostat (dashed orange line) rose exponentially, as in the two-process model. However, unlike in the two-process model, the lower asymptote increased. Then, when the individual was allowed 8-h TIB (three nights, R1–R3), unlike the two-process model, the lower asymptote started to decrease and the homeostat decreased more rapidly than in the chronic sleep restriction (CSR) case. Because the lower asymptote rose to a greater value in the CSR case (panel A) than in the TSD case (panel B), the performance recovery during R1after CSR ($S_{max,A}$–$S_{rec,A}$) was less than that during R1 after TSD ($S_{max,B}$–$S_{rec,B}$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(proportional to the sample variance). This adjusted R^2 takes into account the increase in r^2 with an increasing number of model parameters p and the difference in the number of data points among data sets. The values of R^2 range between $-\infty$ and +1, with +1 indicating zero residual error and negative R^2 indicating that the mean of the data is a better fit than the model.

We also used the explained variance (EV) to describe the variance that the model has in common with the data. EV is defined as the ratio of the variance in the model fits to the variance in the observed data (Van Dongen, 2004), that is $EV = Var(Fit)/Var(Data)$.

2.6. Study data

We used psychomotor vigilance task (PVT) data, a well-validated outcome measure of neurocognitive performance (Belenky et al., 2003; Balkin et al., 2000; Dorrian et al., 2005), to evaluate the proposed unified model. The PVT is a widely used vigilance task in which volunteers respond to a randomly presented visual stimulus approximately 100 times within a test session. The data sets were derived from two versions of the test: (1) a 10-min PVT administered using a small hand-held device (PVT-192—Ambulatory Monitoring Inc., Ardsley, NY), with an inter-stimulus interval varying from 2 to 10 s in 2-s increments (Dinges and Powell, 1985) and (2) a 5-min PVT, administered on a personal digital assistant, with an inter-stimulus interval varying from 1 to 5 s in 1-s increments (Thorne et al., 2005). Performance was quantified by determining the number of lapses (i.e., response times longer than 500 ms); more lapses indicated greater impairment. We analyzed three data sets.

Study A: sleep dose–response study (Belenky et al., 2003; Balkin et al., 2000). PVT performance was measured in a controlled laboratory study of 66 healthy subjects who were assigned to seven nights of 3 h ($n=18$), 5 h ($n=16$), 7 h ($n=16$), or 9 h ($n=16$) of TIB (sleep restriction phase) followed by 3 days of 8-h TIB (recovery phase). Before the sleep restriction phase, all subjects were allowed three nights of 8-h TIB. For all phases, wakeup time was fixed at 07:00 h. A 10-min PVT was administered 4 times per day (09:00, 12:00, 15:00, and 21:00 h) along with other tests reported by Balkin et al. (2000). Subjects in the 3- and 5-h TIB groups performed additional PVT sessions (at 00:00 h for both groups and again at 02:00 h for the 3-h TIB group) during their additional time awake.

Study B: trait-identification study (Rupp et al., 2012). Nineteen healthy adults underwent two sleep-loss challenges separated by 2–4 weeks: (1) TSD, consisting of approximately 63 h (2 nights) of continuous wakefulness, and (2) CSR, consisting of seven nights of 3-h TIB. Both challenges were preceded by seven in-laboratory nights with 10-h TIB and followed by three nights with 8-h TIB (recovery). During the entire wake period, 10-min PVTs were administered every 2 h. For both TSD and CSR, wakeup times were fixed at 07:00 h.

Study C: sleep-banking study (Rupp et al., 2009). Twenty-four subjects were randomly assigned to one of two groups (12 subjects each), sleep banking (10-h TIB) or habitual sleep [mean: 7.09-h TIB, standard deviation (SD): 0.70-h TIB] for seven nights, followed by 1 baseline night (10-h or habitual TIB, respectively), seven sleep restriction nights (3-h TIB), and 5 recovery sleep nights (8-h TIB). A 5-min PVT was administered hourly on a personal digital assistant between 08:00 and 18:00 h during all in-laboratory phases of the study. For both groups, wakeup times were fixed at 07:00 h.

The above laboratory studies were approved by the Walter Reed Army Institute of Research Human Use Review Committee (Silver Spring, Maryland) and the United States (US) Army Medical Research and Materiel Command Human Subjects Research Review Board (Fort Detrick, Maryland), and were performed in accordance with the ethical standards of the 1964 Declaration of Helsinki. Written informed consent was obtained from all subjects before participation.

3. Results

3.1. Simulations

We first used data from *Study A* (sleep dose–response study) to assess the fits of the unified model to population-average PVT lapse data under different sleep-restriction scenarios. Next, using *Study B* (trait-identification study), we cross-validated the model by predicting TSD performance impairment using parameter estimates obtained from fitting CSR PVT lapse data, and vice versa. Finally, we used the parameters obtained from *Study B* to predict PVT lapse data of habitual versus sleep-banked subjects in *Study C* (sleep-banking study) and to illustrate the ability of the unified model to accurately describe performance of subjects with different levels of initial sleep debt.

3.2. Fitting models using the sleep dose–response study (Study A)

We fitted Borbély's two-process model, McCauley et al.'s state-space model, and the unified model to the data from *Study A* and calculated the goodness-of-fit for each model. We estimated one set of parameter values for each model using the population-average PVT lapse data based on the aggregate of all four groups of subjects (233 data points) in a least-squares regression.

Fig. 2 shows the population-average lapse data of the four groups of subjects across seven nights of sleep restriction (3-, 5-, 7-, and 9-h TIB), three nights of recovery sleep (8-h TIB), and the corresponding unified model fits. Table 1 shows the estimates of parameters for each of the three models. Fig. 2 shows that, for each level of sleep restriction, observed performance impairment stabilized in a dose-dependent manner; that is, the performance of subjects who obtained 3-h TIB stabilized at a deteriorated level (i.e., higher number of lapses) compared with the daily mean performance of subjects receiving 5-, 7-, or 9-h TIB. The unified model captured the dose-dependence effect of TIB on performance and fitted the data reasonably well (RMSE = 2.18 lapses, Table 1).

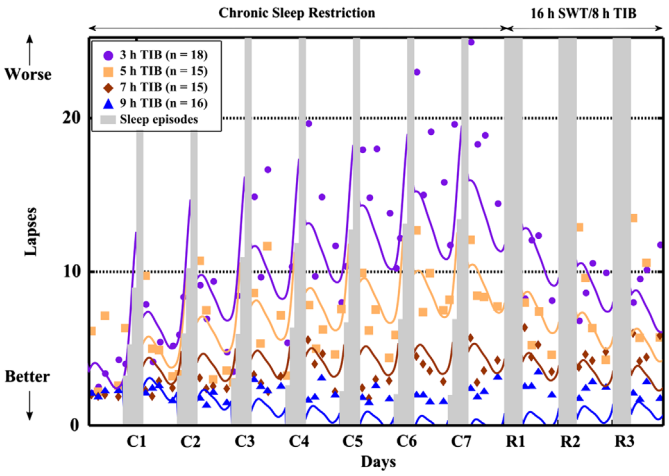

Fig. 2. Population-average PVT lapse data and the fits of the unified model on 3-, 5-, 7-, and 9-h time in bed (TIB) groups for seven nights of chronic sleep restriction (C1–C7), followed by three recovery nights (R1–R3) with 16-h scheduled wake time (SWT) and 8-h TIB.

Table 1
Comparison of the parameter values and goodness-of-fits for the three models for *Study A* (sleep dose–response study). Both the state-space model and the unified model have much better goodness-of-fits than the two-process model as measured by RMSE. However, the state-space model has four more parameters than the unified model and thus the goodness-of-fit, as measured by R^2, is 0.73 for the unified model as compared to 0.47 for the state-space model. RMSE, root mean squared error; R^2, adjusted coefficient of determination; EV, explained variance; τ_s, time constant of the decreasing exponential function during sleep; τ_w, time constant of the increasing saturating exponential function during wakefulness; $L(0)$, lower asymptote at time zero; U, constant value of the upper asymptote; $P(0)$, PVT lapse count at time zero; φ, circadian phase; κ, circadian amplitude; μ, circadian constant; τ_{LA}, lower asymptote time constant. The $\mathbf{A}$ (α_{11}, α_{12}, α_{21}, and α_{22}) and $\mathbf{\Sigma}$ (σ_{11}, σ_{12}, σ_{21}, and σ_{22}) matrices modulate the homeostat and asymptotes during wakefulness and sleep, respectively, with σ_{21} and α_{21} set to 0.00 h.

Model	Parameter estimates					Goodness of fit		
	Parameters	Sleep parameters	Wake parameters	Initial conditions, lapses	Circadian parameters	RMSE, lapses	R^2	EV
Two-process model	7	τ_s=4.51 h	τ_w=36.75 h	$L(0)$=0.00 U=24.30 $S(0)$=0.00	φ=5.47 h κ=4.36 lapses	3.48	0.04	0.54
State-space model	12	σ_{11}=−1.08 h^{-1} σ_{12}=0.44 h^{-1} σ_{22}=−0.09 h^{-1}	α_{11}=−0.03 h^{-1} α_{12}=0.00 h^{-1} α_{22}=0.01 h^{-1}	$L(0)$=5.92 U=36.45 $P(0)$=3.76	φ=12.56 h κ=0.40 lapses μ=0.24 lapses	2.14	0.47	0.56
Unified model	8	τ_s=1.00 h τ_{LA}=4.06 days	τ_w=40.00 h	$L(0)$=0.00 U=24.12 $S(0)$=0.00	φ=2.02 h κ=4.13 lapses	2.18	0.73	0.65

The rate of stabilization was captured by the lower asymptote time constant τ_{LA} in Eq. (5a). τ_{LA} Represents the time that would be required for the lower asymptote to reach 1—1/e≅63.2% of its final value in the case of TSD. Under a TSD scenario, the lower asymptote would reach 63.2% of its final value in 4.06 days (Table 1). Furthermore, it would take 12.18 days ($3\tau_{LA}$) for the lower asymptote to reach 95% of its final value. The unified model also provided a good qualitative description of the experimental observations on the recovery days. The final sleep-debt value after seven nights monotonically increased as TIB decreased; thus, the model captured the observation that groups with less TIB during the sleep restriction phase subsequently displayed less complete recoveries following three nights of 8-h TIB (e.g., the 5-h TIB group displayed a less complete recovery than the 7-h TIB group).

As listed in Table 1, $L(0)$, the value of the lower asymptote at the beginning of the experiment, was estimated to be 0.00 for both the unified model and the two-process model, suggesting that initially subjects carried no sleep debt. In addition, the homeostatic wake parameter and the circadian parameters of the unified model were in the same range as those of the two-process model. These observations were expected because both models use the same equation (Eq. (1a)) to describe the homeostatic process during wakefulness. For the state-space model, the parameter estimates were similar to those obtained in McCauley et al. (2009).

Table 1 also lists the goodness of fit metrics for the three models. The RMSE of the population-average fit obtained using the two-process model was 3.48 lapses; indicating a ~50% larger error when compared with the unified model (2.18 lapses) and the state-space model (2.14 lapses). Taking into account the number of parameters in each model, the adjusted R^2 suggests that the unified model performed ~20% better than the state-space model and considerably better than the two-process model. The unified model also explained 65% of the variance in the population-average data and fitted the data better than the original two-process model (EV=54%) and the state-space model (EV=56%).

The first PVT session of each waking period might be confounded by sleep inertia, i.e., the transient performance impairment frequently observed immediately after awakening (McCauley et al., 2009). The effects of sleep inertia fall outside the predictive capability of the three models. However, we chose not to exclude these data points when analyzing *Study A* because only four to six PVT sessions were available for each day. As a result, the EV was smaller than it would have been had we omitted the first PVT session in each day.

3.3. Estimating performance in the trait-identification study (Study B)

Using data from *Study B*, we cross-validated the models on the TSD and CSR PVT data. For each model, we first estimated the parameters using the population-average PVT lapse data from the CSR scenario (sleep restriction and recovery phases) in a least-squares regression. Using these parameter estimates, we then predicted the population-average PVT lapse count for the same group of subjects undergoing TSD. In the second part of the analysis, we estimated model parameters using the population-average PVT lapse data from the TSD scenario (sleep deprivation and recovery phases) and then predicted the population-average PVT lapse count for the same group of subjects undergoing CSR. In the CSR scenario, we used 10 PVT sessions performed bi-hourly from 09:00 to 03:00 (we excluded the first PVT session at 07:00 to exclude possible sleep inertia).

Fig. 3A and B shows the population-average PVT lapse data, the two-process model fits, and the corresponding predictions for the TSD and CSR scenarios, respectively. Table 2 (top row) lists the sleep homeostatic parameters estimated from the population-average CSR and TSD PVT lapse data and the goodness-of-fit metrics of the fits and predictions for the two-process model. Parameters estimated by separately fitting each of the CSR and TSD data sets were similar (e.g., τ_s=3.97 for CSR and 4.29 h for TSD). For both the TSD and CSR scenarios, errors in predictions from the two-process model (RMSE=2.98 and 2.61 lapses, respectively, Table 2) were up to approximately 50% greater than those from fitting (RMSE=1.97 and 2.05 lapses, respectively, Table 2). While this model fitted the TSD data reasonably well (Fig. 3A), the fits and predictions on the CSR data stabilized by the third day of sleep restriction whereas the observed mean PVT lapse count continued to increase (Fig. 3B). The model also predicted a rapid recovery, which did not match the observed data (Fig. 3B).

Fig. 3C and D shows the population-average PVT lapse data, the state-space model fits, and the corresponding predictions for the TSD and CSR scenarios, respectively. Table 2 (middle row) lists the sleep homeostatic parameters estimated from population-average CSR and TSD PVT lapse data and goodness-of-fit metrics of the fits and predictions for the state-space model. With parameters obtained from fitting the state-space model to the CSR scenario, the model predicted the TSD lapse data reasonably well (Fig. 3C), and the cross-validation RMSE was 3.08 lapses. However, some of the parameters estimated by separately fitting each of the CSR and

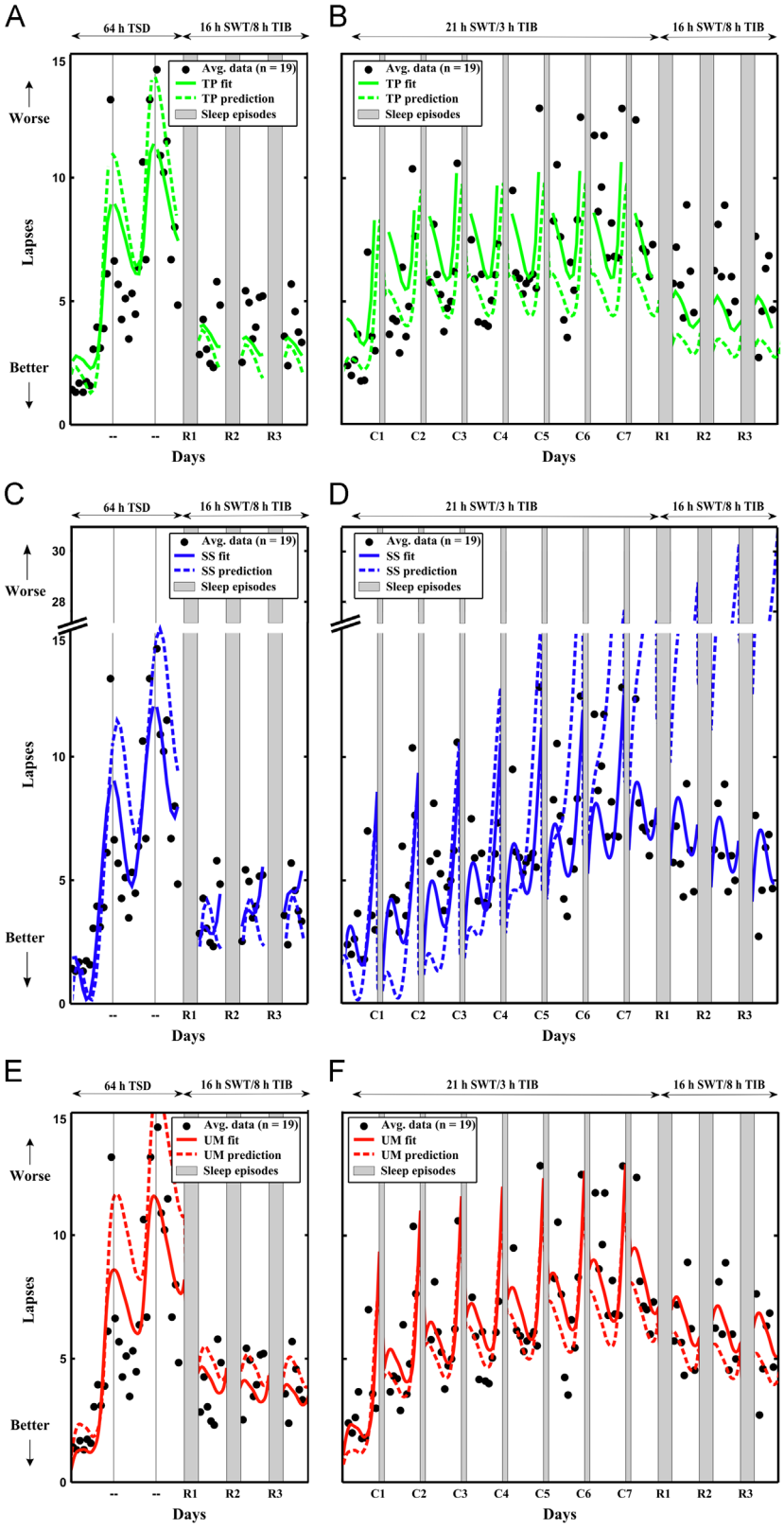

Fig. 3. Population-average PVT lapses and the corresponding fits and predictions. A: Population-average PVT lapses for the for the 64-h total sleep deprivation (TSD) scenario, the fit of the two-process (TP) model, and the cross-validated predictions made using the TP model with parameters estimated from the 21-h scheduled wake time (SWT)/3-h time in bed (TIB) chronic sleep restriction (CSR) scenario; B: population-average PVT lapses for the for the 21-h SWT/3-h TIB CSR scenario, the fit of the TP model, and the cross-validated predictions made using the TP model with parameters estimated from the 64-h TSD scenario; C: same as panel A for the state-space (SS) model; *D*: same as panel B for the SS model; E: same as panel A for the unified model (UM); F: same as panel B for the UM. The cross-validation predictions on the CSR data showed that the TP model (green line) predicted rapid stabilization across days to occur in the CSR scenario, along with a rapid recovery process (panel B), which did not match the observations. The SS model (blue line) predicted that daily mean PVT lapse count continued to degrade in the CSR scenario, suggesting that a daily 3-h TIB did not stabilize the daily mean PVT lapse count across multiple days of sleep restriction (panel D). In contrast, the UM (red line) predicted a gradual stabilization of the daily mean PVT lapse count in the CSR scenario at a level better than that of the TSD scenario (panel F) and fit the data well. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Comparison of the homeostatic parameters in the three models based on cross-validations on TSD PVT lapse data using parameter estimates obtained from fitting CSR PVT lapse data, and vice versa, in *Study B* (trait-identification study). The gray areas indicate fits on CSR data and predictions on TSD data, while the white areas indicate fits on TSD data and predictions on CSR data. Both the unified model's fit on CSR data and prediction on TSD data have smaller RMSE and the larger R^2 than the other two models. The goodness-of-fit of the unified model on TSD data is slightly worse than that of the state-space model; however, the prediction on CSR data from the TSD model parameters is much better. TSD, total sleep deprivation; CSR, chronic sleep restriction; R^2, adjusted coefficient of at time zero; U, constant value of the upper asymptote; τ_s, time constant of the decreasing exponential function during sleep; τ_w, time constant of the increasing saturating exponential function during wakefulness; τ_{LA}, lower asymptote time constant. The $\underline{\mathbf{A}}$ (α_{11}, α_{12}, α_{21}, and α_{22}) and $\underline{\boldsymbol{\Sigma}}$ (σ_{11}, σ_{12}, σ_{21}, and σ_{22}) matrices modulate the homeostat and asymptotes during wakefulness and sleep, respectively, with σ_{21} and α_{21} set to 0.00 h.

Model	Data set	Fits						Predictions	
		Sleep parameters	Wake parameters	$L(0)$, lapses	U, lapses	RMSE, lapses	R^2	RMSE, lapses	R^2
Two-process model	CSR	τ_s=3.97 h	τ_w=20.53 h	0.00	13.15	2.06	0.30		
	TSD							2.99	0.14
	TSD	τ_s=4.29 h	τ_w=23.00 h	0.00	10.42	1.98	0.63		
	CSR							2.61	−0.14
State-space model	CSR	σ_{11}=−0.65 h^{-1} σ_{12}=0.28 h^{-1} σ_{22}=−0.08 h^{-1}	α_{11}=−0.05 h^{-1} α_{12}=0.01 h^{-1} α_{22}=0.01 h^{-1}	−2.65	4.36	1.87	0.37		
	TSD							3.08	−0.03
	TSD	σ_{11}=−15.65 h^{-1} σ_{12}=0.26 h^{-1} σ_{22}=−0.02 h^{-1}	α_{11}=−0.07 h^{-1} α_{12}=0.00 h^{-1} α_{22}=0.02 h^{-1}	10.01	65.52	1.68	0.69		
	CSR							8.08	−10.81
Unified model	CSR	τ_s=2.46 h τ_{LA}=9.90 days	τ_w=27.69 h	−2.50	15.73	1.68	0.52		
	TSD							2.86	0.19
	TSD	τ_s=3.69 h τ_{LA}=6.73 days	τ_w=23.72 h	−2.31	10.91	1.86	0.66		
	CSR							1.99	0.33

TSD data sets were considerable dissimilar (e.g., σ_{11}=−0.65 h for CSR and σ_{11}=−15.65 h for TSD). Parameters estimated from fitting the TSD scenario predicted that daily mean PVT lapse count would continuously increase without limit with days of sleep restriction (Fig. 3*D*). The RMSE of the CSR predictions was 8.08 lapses. Moreover, the bifurcation threshold $T\sigma_{22}/(\sigma_{22}-\alpha_{22})$ calculated using the TSD parameters was 12 h, suggesting that daily mean PVT lapse count would stabilize only if the subjects receive more than 12-h TIB and would continue to degrade otherwise. It is possible that this drastic deviation from the observed data may have arisen from the insufficiency of data points associated with 3 days of TSD followed by 3 days of recovery to estimate the 12 model parameters. However, having limited data available for parameter estimation may be typical of potential study environments, limiting the applicability of the state-space model.

Fig. 3E and F shows the population-average PVT lapse data, the unified model fits, and the corresponding predictions for the TSD and CSR scenarios, respectively. When fitting the TSD and CSR data from *Study B*, we used EEG δ power data to set an upper bound of 240 h on the value of τ_{LA}. Table 2 (bottom row) lists the sleep homeostatic parameters estimated from population-average CSR and TSD PVT lapse counts and goodness-of-fit metrics of the fits and predictions for the unified model. The cross-validation RMSE for the unified model fitted to TSD data to predict the CSR data (1.99 lapses) was only slightly greater than the error when fitting the model to the CSR data itself (1.68 lapses). The cross-validation RMSE when we used the model fitted to CSR data to predict the TSD data (2.86 lapses) was 54% greater than the RMSE when fitting the model to the TSD data itself (1.86 lapses), but still less than the cross-validation RMSEs for the other two models. Also, $L(0)$ was less than zero for both the TSD and CSR scenarios (−2.31 and −2.50 lapses, respectively), indicating that the subjects were sleep banked initially. The time constant τ_{LA} indicated that, for the CSR data, the lower asymptote reached approximately 63.2% of its final value in 9.90 days and that, for TSD data, it would have taken 6.73 days. The wakefulness time constant, τ_w, differed between the TSD and CSR scenarios (23.72 and 27.69 h, respectively). However, the similarity of the errors in the fits and cross-validations suggests that the model fit is fairly insensitive to τ_w. Fig. 3E and F shows the population-average PVT lapse data, the unified model fits, and the corresponding predictions for the TSD and CSR scenarios, respectively. The predictions for both TSD and CSR matched the PVT lapse data well. During CSR, the predictions gradually stabilized at a level better (i.e., lower number of PVT lapses) than that of TSD data (Fig. 3E) and the accuracy of the predictions was similar to that of the fitting. The model also captured the slower recovery process after nights of 8-h TIB following CSR (Fig. 3F), as compared to the faster recovery process after TSD (Fig. 3E). This dissimilarity in the rate of recovery process can be explained by the differences in sleep debt and, thus, differences in the lower asymptote values after 7 days of CSR and 64 h of TSD in the unified model (also shown in Fig. 1).

3.4. Describing the effects of sleep/wake history in the sleep-banking study (Study C)

We used the model with parameters estimated from the CSR data in *Study B* (Table 2) to predict the population-average PVT lapse data of habitual and sleep-banked subjects in *Study C*. This allowed us to test whether the models captured the effects of different sleep/wake history on PVT lapse count.

The model parameters obtained by fitting the CSR data in Table 2 were from subjects who underwent sleep banking (10-h TIB for 7 days before the sleep restriction). Therefore, we used models based on these parameters to predict the population-average PVT lapse count of the sleep-banked group in *Study C*. In the two-process model, the lower asymptote remained at zero regardless of prior sleep-banking conditions. Thus, we used the initial conditions $L(0)$=0.00 lapses and U=13.15 lapses (Table 2) to predict both habitual and sleep-banked groups. In the state-space

model, because the difference between the asymptotes is likely to be different for each dataset, we used the least-squares method to re-estimate the initial values of the lower asymptote $L(0)$ and the upper asymptote U for both the groups while holding the other parameters constant (Table 2, CSR parameters). For the unified model, Eqs. (5b) and (5c) define the relationship between prior sleep debt and $L(0)$. Therefore, we calculated the initial sleep debt of the habitual-sleep group to be 0.06 Eqs. (5a) and (5b) and the lower asymptote value after 7 days of habitual sleep to be 0.81 lapses [$Debt(0) \times U$]. Thus, the initial conditions $L(0)=0.81$ lapses and $U=10.98$ lapses were used to predict the PVT lapse count of the habitual-sleep group and the initial conditions $L(0)=-2.50$ lapses and $U(0)=10.98$ lapses (Table 2, CSR parameters) were used to predict the PVT lapse count of the sleep-banked group, while all other parameters remained the same.

Table 3
Comparison of the PVT lapse count predictions and initial conditions generated by the three models for the habitual and sleep-banked groups in *Study C* (sleep-banking study). The model parameters were estimated from CSR data from Study B. The quality of the unified model prediction of the habitual group data in Study C is better than the predictions of the other two models, while, for the sleep-banked group data in Study C, the quality of the unified model prediction is better than that of the state-space model and similar to that of the two-process model. R^2, adjusted coefficient of determination; RMSE, root mean squared error; $L(0)$, lower asymptote at time zero; U, constant value of the upper asymptote.

Model	Prediction of the habitual group				Prediction of the sleep-banked group			
	R^2	RMSE, lapses	$L(0)$, lapses	U, lapses	R^2	RMSE, lapses	$L(0)$, lapses	$U(0)$, lapses
Two-process model	0.01	2.99	0.00	13.15	0.24	3.23	0.00	13.15
State-space model	0.03	2.89	0.86[a]	8.44[a]	0.09	3.48	7.01[a]	8.06[a]
Unified model	0.17	2.72	0.81[b]	15.73	0.23	3.27	−2.50	15.73

[a] Re-estimated $L(0)$ and $U(0)$ for both sleep groups.
[b] Calculated $L(0)$ for the habitual sleep group.

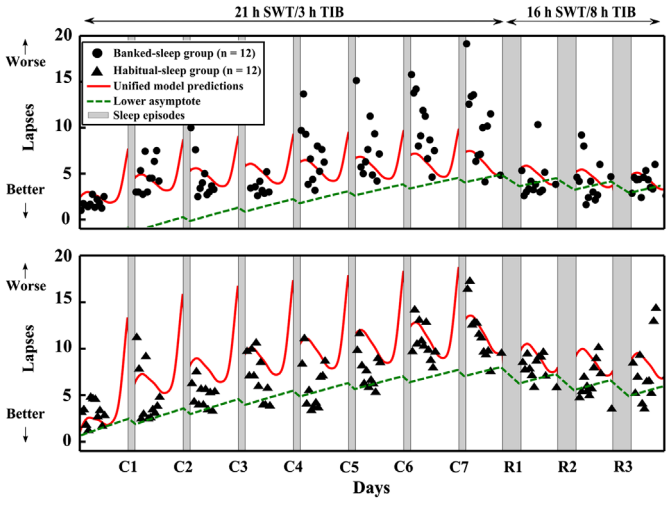

Fig. 4. Population-average performance lapses, the corresponding unified model predictions, and the lower asymptotes of the model. Model parameters were estimated from the chronic sleep restriction data in *Study B*, and performance predictions were made for habitual subjects (*lower panel*) and sleep-banked subjects (*top panel*) in *Study C*. For the unified model, the initial value of the lower asymptote was re-calculated for the habitual sleep group. A 5-min psychomotor vigilance test was used in this study, potentially producing additional noise in the data.

Table 3 lists the initial conditions and goodness-of-fit metrics of the predictions using each of the three models, and Fig. 4 shows the corresponding unified model predictions for both the habitual and sleep-banked groups. The unified model described the effects of habitual sleep and sleep banking as well as or better than the two-process model and the state-space model. For the habitual sleep scenario, the RMSE for the unified model prediction (2.72 lapses) was smaller than the RMSEs for both the two-process model (2.99 lapses) and the state-space model (2.89 lapses). For the sleep banking scenario, the RMSE for the unified model prediction (3.27 lapses) was only slightly greater than the RMSE for the two-process model (3.23 lapses) and smaller than the RMSE for the state-space model (3.48 lapses). Similarly, for the habitual sleep scenario, the unified model R^2 value (0.17) was larger than the R^2 values for both the two-process model (0.01) and the state-space model (0.03). For the sleep banking scenario, the unified model R^2 value (0.23) was slightly less than the R^2 value for the two-process model (0.24) but greater than the R^2 value for the state-space model (0.09).

In addition, we observed that the initial difference between the lower asymptote values of the habitual and sleep-banked subjects was 3.31 [0.81−(−2.50)] lapses and the final difference between the lower asymptote values of the habitual and sleep-banked subjects after 7 days of sleep restriction was 1.55 (6.00–4.45) lapses (Fig. 4). This difference decreased exponentially with days of sleep restriction, indicating that decreases in PVT lapse count caused by sleep banking decreases exponentially with days of sleep restriction.

3.5. *Describing the effects of different PVT measures on fitting and prediction in the trait-identification study (Study B)*

To determine whether our results were affected by the choice of lapse count as a measure of PVT performance, we repeated the cross-validation analysis in Section 3.3 on the unified model, using five different measures of TSD and CSR data from *Study B*: lapses, median response time (RT), mean RT, mean reciprocal response time (RRT), and response time divergence (RTD) (Rajaraman et al., 2012). Table 4 lists the sleep homeostatic parameters estimated from population-average CSR and TSD data and goodness-of-fit metrics of the fits and predictions for the unified model for each of the five PVT measures. To perform the fits on each metric, we constrained τ_s to be not less than 2 h, τ_{LA} to be not greater than 10 days, and θ to lie between 05:00 and 07:00 so that the optimization would generate parameter estimates within expected physiological ranges.

The goodness-of-fit, as measured by R^2, was approximately equal for both TSD and CSR scenarios for lapses, mean RRT, and RTD, slightly lower for mean RT (TSD, 0.57; CSR, 0.40), and lowest for median RT (TSD, 0.34; CSR, 0.31). The cross-validation prediction was highest for mean RRT (TSD, 0.31; CSR, 0.28), similar for lapses, mean RT, and RTD, and lowest for median RT (TSD, 0.14; CSR, 0.13). For all five metrics, the value of R^2 for the fit on TSD data was greater than that for the fit on CSR data. Furthermore, for all metrics except mean RT, the value of R^2 for the cross-validation onto TSD data was greater than that for the cross-validation onto CSR data. Conversely, for all five metrics, the RMSE values for both the fits and cross-validation predictions of TSD data were consistently greater than the corresponding values for the fits and predictions of CSR data; the increased RMSE values for TSD are due to greater variation in the raw data (as shown in Fig. 3).

The estimate of the wake time constant τ_w was similar across all fits, ranging from a low of 20.5 h for TSD using median RT to a high of 32.3 h for CSR using RTD. The estimates of the sleep parameters τ_s and τ_{LA} varied significantly and frequently reached their physiological limits, with τ_s estimated at its lower bound of

Table 4
Comparison of the homeostatic parameters in the unified model calculated using five different PVT measures based on cross-validations on TSD data using parameter estimates obtained from fitting CSR data, and vice versa, in *Study B* (trait-identification study). The gray areas indicate fits on CSR data and predictions of CSR data from models derived using TSD data, while the white areas indicate fits on TSD data and predictions of TSD data from models derived using CSR data. For each measure, the goodness-of-fit on CSR data is better than the goodness-of-fit on TSD. Similarly, for all measures but mean response time, the quality of prediction is better onto CSR than onto TSD. The quality of the fits and predictions are fairly consistent across all measures except for median response time, where the quality is not as good. TSD, total sleep deprivation; CSR, chronic sleep restriction; R^2, adjusted coefficient of determination; RMSE, root mean squared error; $L(0)$, lower asymptote at time zero; U, constant value of the upper asymptote; τ_s, time constant of the decreasing exponential function during sleep; τ_w, time constant of the increasing saturating exponential function during wakefulness; τ_{LA}, lower asymptote time constant; AU, arbitrary units.

Measure	Fits								Predictions	
	Data Set	τ_s (h)	τ_{LA} (days)	τ_w (h)	$L(0)$	U	RMSE	R^2	RMSE	R^2
Lapses	CSR	2.46	9.90	27.7	−2.50 lapses	15.7 lapses	1.68 lapses	0.52	1.99 lapses	0.19
	TSD	3.69	6.73	23.7	−2.31 lapses	10.9 lapses	1.86 lapses	0.66	2.86 lapses	0.33
Median RT	CSR	2.00	3.52	23.7	193 ms	296 ms	15.3 ms	0.31	17.3 ms	0.13
	TSD	3.61	9.99	20.5	207 ms	302 ms	23.6 ms	0.34	26.9 ms	0.14
Mean RT	CSR	2.00	2.92	28.9	204 ms	329 ms	18.2 ms	0.40	20.0 ms	0.27
	TSD	2.00	10.0	25.9	217 ms	349 ms	22.6 ms	0.57	30.4 ms	0.22
Mean RRT	CSR	2.00	2.45	27.5	4.61 ms^{-1}	3.62 ms^{-1}	0.13 ms^{-1}	0.50	0.15 ms^{-1}	0.28
	TSD	2.00	9.99	24.5	4.49 ms^{-1}	3.55 ms^{-1}	0.15 ms^{-1}	0.63	0.20 ms^{-1}	0.31
RTD	CSR	2.20	2.95	32.3	−28.6 AU	27.3 AU	6.02 AU	0.60	7.77 AU	0.22
	TSD	2.20	10.0	22.6	−22.0 AU	27.8 AU	7.66 AU	0.65	11.5 AU	0.32

2 h in six cases and τ_{LA} estimated near its upper bound of 10 days in five cases.

4. Discussion

In this study, we present a unified model that incorporates sleep debt to quantify performance impairment during both CSR and TSD. This model 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, total sleep deprivation to longer periods of sleep restriction (hence, the term “unified”). 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 model as TIB approaches 0 h (i.e., under conditions of acute TSD).

Results from two seminal studies highlighted the need for a new model to describe the dynamics of neurocognitive performance impairment during CSR. In Belenky et al. (2003) and Van Dongen et al. (2003), it was reported that the buildup of cumulative sleep debt occurs in a dose-dependent manner. That is, the more sleep is restricted (both in terms of nightly sleep amounts and in terms of number of nights of restricted sleep), the more next-day neurocognitive performance is impaired (Van Dongen et al., 2003; Alhola and Polo-Kantola, 2007). When sleep was restricted to 5- or 7-h TIB, daily mean performance stabilized at a deteriorated level equivalent to 33% of the performance level seen when subjects obtained their habitual nightly sleep amounts (Belenky et al., 2003; Dinges et al., 1997). Further, Van Dongen et al. (2003) noted that performance continued to degrade (i.e., did not stabilize) when sleep was restricted to 4 h per day. They also observed that recovery of neurocognitive functioning after CSR was not as rapid as that seen after TSD, suggesting that CSR induces relatively long-term, slow-recovering (i.e., long time-constant) changes in brain physiology that affect alertness and performance (Belenky et al., 2003; Johnson et al., 2004; Alhola and Polo-Kantola, 2007).

The classical two-process model does not describe this cumulative sleep debt observed during CSR (Van Dongen et al., 2003; Carskadon and Dement, 1981) or its dissipation following CSR, leading to discrepancies between predictions based on the two-process model and observed PVT lapse data collected during controlled CSR studies (Van Dongen et al., 2003; Achermann, 2004). The unified model extends the theoretical framework laid down by Borbély's classical two-process model by including an additional sleep-debt component, $Debt(t)$, to account for sleep/wake history. The unified model represents sleep debt accumulation and restoration with a recursive filter that gradually diminishes the effect of sleep loss and sleep extensions as they recede further into the past. The recursive filter in the unified model in an exponential filter that performs a similar function to the moving-average filter in the FAID model (Dawson and Fletcher, 2001) and also incorporates the possibility of sleep banking.

We assessed the performance of our proposed model on three data sets: a sleep dose–response study, a trait-identification study, and a sleep-banking study. Using PVT lapse data as a well-validated outcome measure of neurocognitive performance (Belenky et al., 2003; Balkin et al., 2000; Dorrian et al., 2005), the modeling results indicate that daily mean performance impairment eventually stabilizes for all cases of sleep restriction. The rate of performance stabilization is not as rapid as in the two-process model and more closely matches the observed data. The results also suggest that sleep debt accumulation is responsible for the incomplete recovery process after three nights of 8-h TIB following CSR. In a case of CSR where an individual is subject to 3-h TIB for seven nights, the sleep debt value is much higher than when the individual is subject to TSD for 64 h. We observed that although the increase in the homeostat S occurred more rapidly in the case of TSD, it increased for only 64 h, limiting the amount of sleep debt to less than the CSR case. This difference in sleep debt explains the slower recovery process following CSR.

Also of interest is whether one's response to TSD serves to reliably predict one's response to CSR. Of the three models analyzed (two-process model, state-space model, and unified model), the unified model parameters estimated by fitting performance during TSD provided sufficient information to accurately

describe the response to chronic sleep restriction. This, however, was not the case for the state-space model. The state-space model predicted that PVT lapses would continually degrade in the 3-h TIB scenario. The parameters estimated from the TSD scenario indicate that the rate of bifurcation is much faster than what was estimated by McCauley et al. (2009). This discrepancy could have occurred because the number of data points collected in 3 days of TSD followed by 3 days of recovery were insufficient to accurately estimate the 12 parameters of the model. However, a lack of sufficient data for accurate parameter estimation may be quite typical of operational environments, limiting the applicability of the state-space model. In contrast, the two-process model predicted rapid stabilization of daily mean PVT lapse count during sleep restriction, a prediction which did not match the data.

The unified model introduces only one parameter, τ_{LA}, that does not appear in the classical two-process model. This one additional parameter is sufficient to improve (often considerably) the goodness-of-fit of the unified model when compared to that of the two-process model, even when evaluating the fit using the parameter-adjusted coefficient of determination R^2, which penalized the unified model for its extra parameter. Because the state-space model has four more parameters than the unified model, the state-space model can sometimes better fit the data (e.g., RMSE of 1.68 lapses vs. 1.86 lapses for the TSD fits in Table 2). However, the additional degrees of freedom introduced by the state-space model more drastically impaired the *predictive* accuracy of this less parsimonious model. This is, in part, because it is harder to accurately estimate the larger number of parameters in the state-space model and less accurate parameter estimates lead to less accurate predictions (e.g., RMSE of 8.08 lapses vs. 1.99 lapses for the CSR predictions in Table 2).

An advantage of the proposed unified model is that 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. Another advantage is that the model captures the effects of sleep banking and the rate at which alertness and performance are restored after recovery sleep. The modulation of the lower asymptote $L(t)$ captures mathematically the findings of Rupp et al. (2009), in which it was reported that the extent to which sleep restriction impairs alertness varies as a function of prior sleep/wake history. The modulation of the lower asymptote also accounts for the diminishing effects of sleep banking across long periods of CSR. Because the influence of the level of sleep debt decreases exponentially with days of sleep restriction, the difference between the lower asymptotes of sleep-banked subjects and non-sleep-banked subjects decreased exponentially as well, reducing the difference from 3.31 to 1.55 lapses after 7 days of 3-h TIB (Fig. 4). Our analysis differs from the original analysis of Rupp et al. (2009), in which a linear model was fitted onto the PVT data, and no PVT data were collected after 18:00. A linear model does not capture the exponential decay of the effects of sleep banking described by the unified model and thus does not extrapolate into the future when the effects of sleep banking are predicted to reduce to zero.

The results from the present analyses indicate that the unified model fits CSR data well. Although we did not consider such scenarios in this study, the unified model can also be applied to studies incorporating sleep shifting, daytime naps, and other irregular sleep patterns. The unified model can be extended to incorporate the decreased restorative effects of sleep opportunities at other circadian phases (Dijk and Czeisler, 1994) by incorporating time of day into the calculation of the value $Loss(t)$ in Eq. (5b). Furthermore, in the case of very long TIB periods, when the duration of sleep may be considerably less than the TIB, the value of $Loss(t)$ can be modified to take into account the actual length of sleep, instead of the length of TIB. Other modifications involve accounting for trait-like individual differences in parameters, such as the sleep requirements for full performance, and providing an estimate of the group variance (Rajaraman et al., 2008, 2009). These modifications would help in building individual-specific models.

We evaluated the ability of the unified model to fit and predict TSD and CSR performance using five different PVT measures and found that the quality of fits and predictions, as measured by RMSE and R^2, was similar across these PVT measures (with the quality being slightly worse for median RT than the other measures, as shown in Table 4). The five PVT measures we considered are highly correlated with each other (Rajaraman et al., 2012) and thus we would expect them to produce a generally similar quality of fits and predictions. For all of the PVT measures, we consistently found that the values of RMSE and R^2 were both greater for TSD data fits than for CSR data fits; the greater RMSE values for TSD data fits are due to greater variability in TSD data (Fig. 3). The same pattern of greater RMSE and R^2 values for TSD data than CSR data was also observed in the cross-validation predictions, with the sole exception being a greater R^2 value for the prediction of CSR performance using mean RT (0.27 CSR; 0.22 TSD).

The sleep parameters in the unified model, τ_{LA} and τ_s, were not consistently estimated across TSD and CSR scenarios (Table 4). These estimates of the lower asymptote time constant τ_{LA} and the sleep time constant τ_s are closely connected and both of these parameters are "invisible" to PVT data in the sense that their direct effect on predicted performance occurs during sleep, when no measurements can be taken. The differences between estimates of τ_{LA} and τ_s between CSR and TSD may be due to the differences between the two phases of *Study C*. During the CSR phase, there were seven nights of 3-h TIB and three nights of 8-h TIB (i.e., 2×10 data points) that were used to estimate τ_{LA} and τ_s. However, in the TSD phase, there were only three nights of 8-h TIB recovery sleep (i.e., 2×3 data points) from which to estimate these parameters. Furthermore, different amounts of TIB (e.g., 3-h TIB vs. 8-h TIB) could result in different estimates of τ_{LA} and τ_s. Determining parameter estimates that are generalizable across different studies for τ_{LA} and τ_s would likely involve incorporating additional physiological measurements, such as EEG δ power, into the estimation procedure.

In summary, this work introduces a new approach to model the neurocognitive performance degradation of sleep-restricted individuals, as measured by PVT lapse data. The unified model seamlessly bridges the continuum from CSR to TSD by modeling TSD as a limiting case of CSR and it naturally converges to the well-established two-process model when TIB approaches zero. Furthermore, this model also explains the slow neurocognitive recovery process after CSR and the improved neurocognitive performance of sleep-banked subjects.

Grants

This work was funded by the Military Operational Medicine Research Program (MOMRP) of the US Army Medical Research and Materiel Command (Ft. Detrick, MD), and by a Grant from US Defense Health Program (Award # 13200) managed by the MOMRP.

Disclosure statement

This was not an industry supported study. The authors have indicated no financial conflicts of interest. 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 US Army or of the US Department of Defense.

Author contributions

P.R., D.T., S.R., and J.R. conceived research; P.R. implemented the model; T.L.R., N.J.W., and T.J.B. provided data for modeling; P.R. and D.T. wrote the paper, which was edited by N.J.W. and J.R.

Acknowledgments

The authors thank Dr. Sridhar Ramakrishnan, Dr. Srinivas Laxminarayan, and Dr. Wei Lu for helpful discussions.

References

Achermann, P., 2004. The two-process model of sleep regulation revisited. Aviat. Space Environ. Med. 75, A37–A43.
Achermann, P., Borbély, A.A., 1992. Combining different models of sleep regulation. J. Sleep Res. 1, 144–147.
Achermann, P., Borbély, A.A., 1994. Simulation of daytime vigilance by the additive interaction of a homeostatic and a circadian process. Biol. Cybern. 71, 115–121.
Alhola, P., Polo-Kantola, P., 2007. Sleep deprivation: impact on cognitive performance. Neuropsychiatr. Dis. Treat. 3, 553–567.
Avinash, D., Crudele, C.P., Amin, D.D., Robinson, B.M., Dinges, D.F., Van Dongen, H.P., 2005. Parameter estimation for a biomathematical model of psychomotor vigilance performanesce under laboratory conditions of chronic sleep restriction. Sleep-Wake Research in the Netherlands, vol. 16. Dutch Society for Sleep-Wake Research, Leiden, The Netherlands, pp. 39–42.
Balkin, T.J., Belenky, G., Hall, S., Williams, J., Wesensten, N.J., Redmond, D.P., Thomas, M.L., Sing, H.C., Thome, D.R., 2000. Effects of Sleep Schedules on Commercial Motor Vehicle Driver Performance. US DOT, National Technical Information Service, Springfield, VA.
Belenky, G., Wesensten, N.J., Thorne, D.R., Thomas, M.L., Sing, H.C., Redmond, D.P., Russo, M.B., Balkin, T.J., 2003. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose–response study. J. Sleep Res. 12, 1–12.
Borbély, A.A., 1982. A two process model of sleep regulation. Hum. Neurobiol. 1, 195–204.
Borbély, A.A., Achermann, P., 1999. Sleep homeostasis and models of sleep regulation. J. Biol. Rhythms 14, 557–568.
Carskadon, M.A., Dement, W.C., 1981. Cumulative effects of sleep restriction on daytime sleepiness. Psychophysiology 18, 107–113.
Daan, S., Beersma, D.G., Borbély, A.A., 1984. Timing of human sleep: recovery process gated by a circadian pacemaker. Am. J. Physiol. Regul. Integr. Comp. Physiol. 246, R161–R183.
Dawson, D., Fletcher, A., 2001. A quantitative model of work-related fatigue: background and definition. Ergonomics 44, 144–163.
Dijk, D.-J., Czeisler, C.A., 1994. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. Neurosci. Lett. 166, 63–68.
Dinges, D.F., Powell, J., 1985. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav. Res. Methods 17, 652–655.
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 h per night. Sleep 20, 267–277.
Dorrian, J., Rogers, N.L., Dinges, D.F., 2005. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In: Kushida, C.A. (Ed.), Sleep Deprivation. Clinical Issues, Pharmacology, and Sleep Loss Effects. Marcel Dekker, New York, pp. 39–70.
Friedl, K.E., Mallis, M.M., Ahlers, S.T., Popkin, S.M., Larkin, W., 2004. Research requirements for operational decision-making using models of fatigue and performance. Aviat. Space Environ. Med. 75, A192–A199.
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. Aviat. Space Environ. Med. 75, A44–A53.
Johnson, M.L., Belenky, G., Redmond, D.P., Thorne, D.R., Williams, J.D., Hursh, S.R., Balkin, T.J., 2004. Modulating the homeostatic process to predict performance during chronic sleep restriction. Aviat. Space Environ. Med. 75, A141–A146.
Klerman, E.B., St-Hilaire, M., 2007. On mathematical modeling of circadian rhythms, performance, and alertness. J. Biol. Rhythms 22, 91–102.
Mallis, M.M., Mejdal, S., Nguyen, T.T., Dinges, D.F., 2004. Summary of the key features of seven biomathematical models of human fatigue and performance. Aviat. Space Environ. Med. 75, A4–A14.
McCauley, P., Kalachev, L.V., Smith, A.D., Belenky, G., Dinges, D.F., Van Dongen, H.P., 2009. A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. J. Theor. Biol. 256, 227–239.
Mollicone, D.J., Van Dongen, H.P., Rogers, N.L., Banks, S., Dinges, D.F., 2010. Time of day effects on neurobehavioral performance during chronic sleep restriction. Aviat. Space Environ. Med. 81, 735–744.
Rajaraman, S., Gribok, A.V., Wesensten, N.J., Balkin, T.J., Reifman, J., 2008. Individualized performance prediction of sleep-deprived individuals with the two-process model. J. Appl. Physiol. 104, 459–468.
Rajaraman, S., Gribok, A.V., Wesensten, N.J., Balkin, T.J., Reifman, J., 2009. An improved methodology for individualized performance prediction of sleep-deprived individuals with the two-process model. Sleep 32, 1377–1392.
Rajaraman, S., Ramakrishnan, S., Thorsley, D., Wesensten, N.J., Balkin, T.J., Reifman, J., 2012. A new metric for quantifying performance impairment on the psychomotor vigilance test. J. Sleep Res. 21, 659–674.
Rupp, T.L., Wesensten, N.J., Bliese, P.D., Balkin, T.J., 2009. Banking sleep: realization of benefits during subsequent sleep restriction and recovery. Sleep 32, 311–321.
Rupp, T.L., Wesensten, N.J., Balkin, T.J., 2012. Trait-like vulnerability to total and partial sleep loss. Sleep 35, 1163–1172.
Rusterholz, T., Durr, R., Achermann, P., 2010. Inter-individual differences in the dynamics of sleep homeostasis. Sleep 33, 491–498.
Thorne, D.R., Johnson, D., Redmond, D.P., Sing, H.C., Belenky, G., Shapiro, J., 2005. The Walter Reed palm-held psychomotor vigilance test. Behav. Res. Methods 37, 111–118.
Van Dongen, H.P., 2004. Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviat. Space Environ. Med. 75, A15–A36.
Van Dongen, H.P., 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, 117–126.