Real-time individualization of the unified model of performance

JIANBO LIU[1], SRIDHAR RAMAKRISHNAN[1], SRINIVAS LAXMINARAYAN[1], 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, US Army Medical Research and Materiel Command, Fort Detrick, MD, USA; [2]Behavioral Biology Branch, Walter Reed Army Institute of Research, Silver Spring, MD, USA

Keywords
Bayesian learning, individualized modelling, performance prediction, recursive parameter estimation

Correspondence
Jaques Reifman, PhD, Senior Research Scientist and Director, Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, ATTN: MCMR-TT, 504 Scott Street, Fort Detrick, MD 21702, USA.
Tel.: 301-619-7915;
fax: 301-619-1983;
e-mail: jaques.reifman.civ@mail.mil

Accepted in revised form 10 March 2017; received 4 January 2017

DOI: 10.1111/jsr.12535

SUMMARY

Existing mathematical models for predicting neurobehavioural performance are not suited for mobile computing platforms because they cannot adapt model parameters automatically in real time to reflect individual differences in the effects of sleep loss. We used an extended Kalman filter to develop a computationally efficient algorithm that continually adapts the parameters of the recently developed Unified Model of Performance (UMP) to an individual. The algorithm accomplishes this in real time as new performance data for the individual become available. We assessed the algorithm's performance by simulating real-time model individualization for 18 subjects subjected to 64 h of total sleep deprivation (TSD) and 7 days of chronic sleep restriction (CSR) with 3 h of time in bed per night, using psychomotor vigilance task (PVT) data collected every 2 h during wakefulness. This UMP individualization process produced parameter estimates that progressively approached the solution produced by a *post-hoc* fitting of model parameters using all data. The minimum number of PVT measurements needed to individualize the model parameters depended upon the type of sleep-loss challenge, with ~30 required for TSD and ~70 for CSR. However, model individualization depended upon the overall duration of data collection, yielding increasingly accurate model parameters with greater number of days. Interestingly, reducing the PVT sampling frequency by a factor of two did not notably hamper model individualization. The proposed algorithm facilitates real-time learning of an individual's trait-like responses to sleep loss and enables the development of individualized performance prediction models for use in a mobile computing platform.

INTRODUCTION

Mathematical models are being used to forecast the effects of sleep/wake cycles on neurobehavioural performance in order to design work schedules that optimize alertness while on duty and minimize the potential for fatigue-induced accidents. To date, such models have been used primarily as off-line planning tools to predict the performance of an 'average' individual (Dawson *et al.*, 2011; Hursh *et al.*, 2004). However, given the large intersubject variability in the response to sleep loss (Van Dongen *et al.*, 2004), implementing these models in mobile computing smartphone devices would allow for individualized model customization and more accurate predictions (Ramakrishnan *et al.*, 2015).

During the last decade, a handful of new models have been proposed to address the shortcomings of previous neurobehavioural performance models (Mallis *et al.*, 2004). Nevertheless, none possesses the mathematical formalism required to individualize performance models in real time across various sleep/wake conditions. For example, some models predict the performance of individuals subjected to TSD, but not of those subjected to CSR (Rajaraman *et al.*, 2008, 2009; Van Dongen *et al.*, 2007). A few are unsuitable for running real-time applications on a mobile platform because they use computationally expensive optimization algorithms for individualization (Van Dongen *et al.*, 2007), whereas others are impractical because they require regular sampling of individual performance data, which

are required for model individualization (Rajaraman *et al.*, 2008, 2009).

Recently, our group developed the UMP, which predicts individual performance on PVT data accurately under conditions ranging from CSR to TSD and 'learns' an individual's trait-like response to sleep loss (Ramakrishnan *et al.*, 2015). However, individualization of such models has thus far been achieved in a *post-hoc* manner by fitting the model parameters to the complete set of PVT data collected during the course of an entire sleep-loss challenge.

In contrast, personal mobile applications capable of predicting an individual's response to sleep loss in real time require algorithms that continually and efficiently adapt model parameters 'on the fly', as each new measure of individual performance becomes available. Here, we attempted to achieve this via an extended Kalman filter algorithm (Arulampalam *et al.*, 2002) that progressively adapts the UMP model parameters within the framework of Bayesian learning (Chen, 2003). This algorithm lends itself to a computationally efficient learning strategy in which model parameters are individualized recursively solely by the most recent PVT measurement, using simple algebraic computations.

By using experimental data from a cross-over study (Rupp *et al.*, 2012) to simulate model individualization in real time, we addressed the following questions. How does the model individualization achieved by the new algorithm compare with results obtained by using the complete set of PVT data to fit the model? How many PVT measurements are needed to individualize the UMP under different sleep-loss challenges? How does the number of daily PVT measurements impact the rate of model individualization?

MATERIALS AND METHODS

Study data

We analysed PVT data from a cross-over design study (Rupp *et al.*, 2012) in which 18 healthy adults (mean age: 28 years; range; 18–39) underwent the following two sleep-loss challenges separated by 2–4 weeks: (1) 64 h of TSD and (2) seven consecutive nights of CSR, consisting of 3 h of time in bed (TIB) per night. During wake periods, 10-min PVT sessions were administered every 2 h [51 and 85 total measurements (*N*) for the TSD and CSR challenges, respectively]. Using response time (RT) data from each of these PVT sessions, we computed the following five performance statistics, the first four after removing outliers (RTs <100 ms or RTs >3000 ms): (1) mean RT, (2) median RT, (3) slowest 10% RT, (4) speed (mean 1/RT) and (5) lapses (number of RTs >500 ms).

Unified Model of Performance (UMP)

The recently developed and validated UMP forms the core of our predictive modelling framework (Rajdev *et al.*, 2013; Ramakrishnan *et al.*, 2015, 2016). It is based on Borbély's classical two-process model of sleep regulation (Borbély and Achermann, 1999), which we extended to account for the effects of both total and partial sleep loss. We achieved this by explicitly considering the amount of sleep debt resulting from known sleep/wake histories, and modulating the recovery capacity during sleep to vary inversely with sleep debt. Table 1 shows the equations governing the UMP [equations (1)–(4)], whose eight unknown parameters must be determined to individualize the model.

Although the time constants τ_w, τ_s and τ_{LA} are important for modelling the exponential rise and decay of sleep pressure and sleep debt, because the UMP output is not as sensitive to these parameters (see Supporting information, Table S1), we fixed them to 18.2, 4.2 h (Borbély and Achermann, 1999) and 7 days (Ramakrishnan *et al.*, 2015), respectively. Therefore, we estimated only five UMP parameters for each subject: $\theta = [U, \kappa, \phi, S_0, L_0]^T$. Here, U and L_0 denote the upper asymptote and the initial state value of the lower asymptote, respectively, of the homeostatic process; κ and ϕ denote the circadian amplitude and phase, respectively; and S_0 denotes the initial homeostatic sleep pressure.

Post-hoc and Bayesian learning for model individualization

Previously, we individualized the UMP by fitting the model parameters θ to the complete set of PVT measurements obtained in the study (Ramakrishnan *et al.*, 2015). In this *post-hoc* approach, the UMP learns an individual's trait-like response to sleep loss *en masse* by minimizing the sum of squared errors between the complete set of PVT measurements and the corresponding predicted performance [equation (5), Table 2]. However, this approach cannot be used for real-time model individualization, because when used with too few measurements it can lead to unreliable estimates of θ and inferior predictions without proper regularization (Seber and Wild, 2003).

We can use Bayesian learning to address this issue [equation (6), Table 2]. Because this method considers the prior knowledge of a group-average model, model individualization becomes more reliable with each new PVT measurement (Olofsen *et al.*, 2004; Rajaraman *et al.*, 2009). With only a few measurements (i.e. when n is small), the solution to equation (6) through non-linear optimization is largely weighted by the first term, leading to individualized models approximating the model of the average individual. However, as n increases, the weighting of the second term increases, leading to individualized models that represent an individual's sleep-loss phenotype. In the extreme case, where $n \to \infty$, the model obtained by optimizing equation (6) converges asymptotically to the best-fitted model obtained by solving equation (5).

Table 1 Biomathematical equations governing the Unified Model of Performance (UMP)

UMP governing equations

Performance impairment:

$$f(t,\theta) = S(t) + \kappa C(t) \quad (1)$$

where θ comprises the eight model parameters of the UMP, with $\theta = [U, \tau_w, \tau_s, \tau_{LA}, \kappa, \phi, S_0, L_0]^T$ as defined below. $S(t)$ and $C(t)$ denote the homeostatic and circadian processes at time t, respectively, and κ represents the circadian amplitude.

Circadian process (C):

$$C(t) = \sum_{j=1}^{5} a_j \sin\left[j\tfrac{2\pi}{\tau}(t+\phi)\right] \quad (2)$$

where a_j, $j = 1, \ldots, 5$, represent the amplitude of the five harmonics ($a_1 = 0.97$, $a_2 = 0.22$, $a_3 = 0.07$, $a_4 = 0.03$, and $a_5 = 0.001$), τ denotes the fundamental period of the circadian clock (~24 h) and ϕ denotes the circadian phase.

Homeostatic process (S):

$$\dot{S}(t) = \begin{cases} 1/\tau_w[U - S(t)] & \text{during wakefulness} \\ -1/\tau_s[S(t) - L(t)] & \text{during sleep} \end{cases} \quad (3)$$

where U and L denote the upper and lower asymptotes, respectively, and τ_w and τ_s denote the wake and sleep time constants of the increasing and decreasing sleep pressure, respectively. $S(0) = S_0$ and $L(0) = L_0$ correspond to the initial state values for S and L.

Lower asymptote (L) of process S:

$$L(t) = \begin{cases} \max\{U - (U - L_0)\exp(-t/\tau_{LA}), -0.11U\} & \text{during wakefulness} \\ \max\{-2U + (2U + L_0)\exp(-t/\tau_{LA}), -0.11U\} & \text{during sleep} \end{cases} \quad (4)$$

where τ_{LA} denotes the time constant of the exponential decay of the effect of sleep history on performance.

Real-time recursive model individualization

A computationally efficient approach that does not require non-linear optimization to individualize the UMP is to approximate the solution of the Bayesian optimization problem in equation (6), using an extended Kalman filter formulation (Arulampalam *et al.*, 2002). We can estimate the model parameters $\hat{\theta}_n$ recursively, at current time t_n, as a function of the previous estimate $\hat{\theta}_{n-1}$, at time t_{n-1} and the current PVT measurement y_n by solving two algebraic equations [equations (7) and (8), Table 2]. The approximate nature of the estimate $\hat{\theta}_n$ stems from the first-order Taylor series expansion used to compute the Jacobian J_n of the non-linear function $f(t_n, \theta)$ in equation (1) (Arulampalam *et al.*, 2002).

To start the recursion, we assume that $\hat{\theta}_0 = \theta_0$ and $\hat{\Sigma}_0 = \Sigma_0$, where θ_0 and Σ_0 denote priors as in equation (6). We used non-linear mixed-effect modeling (Olofsen *et al.*, 2004) to estimate the group-average model parameters θ_0 and the corresponding variance–covariance matrix Σ_0 for the UMP. To ensure that θ_0 and Σ_0 did not contain information about the subject whose performance we sought to predict, we estimated them using an independent data set of 20 subjects (hourly 10-min PVT sessions during wakefulness) who underwent 5 days of 5 h TIB, followed by 3 days of 8 h TIB (So *et al.*, 2016).

We also computed confidence intervals (CIs) for the UMP parameters and prediction intervals for the UMP outputs at each time t_n (Seber and Wild, 2003). We assumed that the model parameters (and outputs) followed a multivariate Gaussian distribution asymptotically with mean $\hat{\theta}_n[f(t, \hat{\theta}_{n-1})]$ and variance–covariance matrix $\hat{\Sigma}_n[J^T\hat{\Sigma}_n J + \sigma^2]$.

Measures of model convergence

To assess the convergence of the parameters estimated by the recursive learning algorithm throughout all 18 subjects, we computed the intraclass correlation coefficient (ICC) between the recursively estimated parameters $\hat{\theta}_n$ after n ($n \leq N$) PVT measurements and the parameters of the best-fitted model θ^* obtained by using the complete set of data ($N = 51$ or 85 for the TSD and CSR challenges, respectively). Higher ICC values indicated greater parameter agreement between the recursively learned and best-fitted models. We used the following ranges (Landis and Koch, 1977) to interpret the ICC values: slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) and almost perfect (0.81–1.00) agreement.

To assess the ability of the algorithm to yield accurate UMP predictions (i.e. outputs) after n PVT measurements, we computed the absolute root mean squared error (RMSE) between the predictions and measurements. In addition, we computed the relative RMSE, defined as the difference in absolute RMSEs between the recursively learned and best-fitted models divided by the absolute RMSE of the best-fitted model:

$$\text{relative RMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}[y_i - f(t_i, \hat{\theta}_n)]^2} - \sqrt{\frac{1}{N}\sum_{i=1}^{N}[y_i - f(t_i, \theta^*)]^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}[y_i - f(t_i, \theta^*)]^2}} \times 100\% \quad (9)$$

Table 2 Individualization of the Unified Model of Performance (UMP)

Post-hoc model individualization:

$$\arg\min_{\theta}\left\{\sum_{i=1}^{N}[y_i - f(t_i,\theta)]^2\right\} \quad (5)$$

where y_i, $i = 1, 2,\ldots, N$, are the *complete* set of N psychomotor vigilance task (PVT) measurements, and $f(t_i, \theta)$, $i = 1, 2,\ldots, N$, are the UMP predicted performance measurements at discrete times t_i, $i = 1, 2,\ldots, N$. The solution of equation (5) leads to the 'best-fitted' model, with optimal parameter θ^*.

Bayesian learning:

$$\arg\min_{\theta}\left\{(\theta-\theta_0)^T\sum_0^{-1}(\theta-\theta_0) + \frac{1}{\sigma^2}\sum_{i=1}^{n}[y_i - f(t_i,\theta)]^2\right\} \quad (6)$$

where the mean parameter θ_0 represents an 'average' individual, Σ_0 represents the prior variance–covariance matrix of the model parameters θ_0, and σ^2 denotes the noise variance in PVT measurements y_i. The solution to equation (6) leads to the individualized model based on an individual's own set of n PVT measurements y_i, with $i = 1, 2,\ldots, n$, up to the current time t_n (where $n \leq N$).

Recursive learning based on an extended Kalman filter:

We can recursively estimate the model parameter $\hat{\theta}_n$, at current time t_n, with $n = 1, 2,\ldots, N$, as a function of the previous estimate $\hat{\theta}_{n-1}$ at time t_{n-1} and the current PVT measurement y_n, by solving the following algebraic equations:

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \frac{\hat{\sum}_{n-1}J_n}{\sigma^2 + J_n^T\hat{\sum}_{n-1}J_n}[y_n - f(t_n,\hat{\theta}_{n-1})] \quad (7)$$

$$\hat{\sum}_n = \left(I - \frac{\hat{\sum}_{n-1}J_nJ_n^T}{\sigma^2 + J_n^T\hat{\sum}_{n-1}J_n}\right)\hat{\sum}_{n-1} \quad (8)$$

where $\hat{\Sigma}_n$ and $\hat{\Sigma}_{n-1}$ denote the estimated variance–covariance matrices of the model parameters at times t_n and t_{n-1}, respectively, $J_n = \partial f(t_n,\theta)/\partial\theta|_{\theta=\hat{\theta}_{n-1}}$ represents the Jacobian of the model output with respect to the model parameters at time t_n, and I represents the identity matrix. To start the recursion, we assume that $\hat{\theta}_0 = \theta_0$ and $\hat{\Sigma}_n = \Sigma_0$, where θ_0 and Σ_0 denote priors as in equation (6).

We computed absolute and relative RMSEs for each subject as well as their values averaged across the 18 subjects.

RESULTS

We simulated real-time performance by sequentially providing each of the N PVT measurements as inputs to the recursive algorithm, updating the UMP model parameters after each measurement, and using the updated model to predict an individual's performance. We first used mean RT statistics to compare these results with those obtained by the best-fitted model.

Convergence of the recursive algorithm

To assess the ability of the UMP to learn an individual's response to sleep loss with the recursive algorithm, we evaluated its temporal convergence to the best-fitted model. Fig. 1 shows the measured PVT performance data for two individuals (top and bottom panels) while undergoing the TSD (left panels) and CSR (right panels) challenges, along with the results for the best-fitted model, group-average model and recursive algorithm applied to the first 15 and 30 measurements (TSD) or 30 and 60 measurements (CSR). For each subject and challenge, the performance trajectories for the recursive algorithm became progressively closer to those of the best-fitted model as the number of measurements n increased. The group-average model overestimated the mean RT for subject 11 consistently, highlighting the benefit of developing individualized models.

Fig. 2 shows the corresponding values of the recursive algorithm estimates for the three most sensitive UMP model parameters, U, κ and ϕ (Rajdev *et al.*, 2013; Ramakrishnan *et al.*, 2015), their associated 95% CIs, the best-fitted model parameters (horizontal red lines), the absolute RMSEs and the relative RMSEs as a function of the number of PVT measurements n. As n increased, each parameter converged toward its counterpart of the best-fitted model and its 95% CI decreased progressively. For both subjects and across the two challenges, the relative RMSEs between the recursive algorithm and the best-fitted model approached zero as n increased, indicating that the differences in the model outputs became increasingly smaller during the course of learning (Fig. 1). The reduction in the relative RMSEs was most pronounced at the beginning of the learning process, when large discrepancies between PVT data and recursive model outputs led to large parameter adjustments in equation (7). For all subjects, the recursively estimated parameters converged progressively toward their respective counterparts of the best-fitted model as more PVT measurements became available (also see Supporting information, Fig. S1 for all subjects).

Number of PVT measurements required for model individualization

We computed the ICCs between the parameters estimated by the recursive algorithm and those obtained by the best-fitted model as a function of the number of PVT measurements (Fig. 3). For TSD, the recursive algorithm required 27 PVT measurements sampled every 2 h for the ICC of the

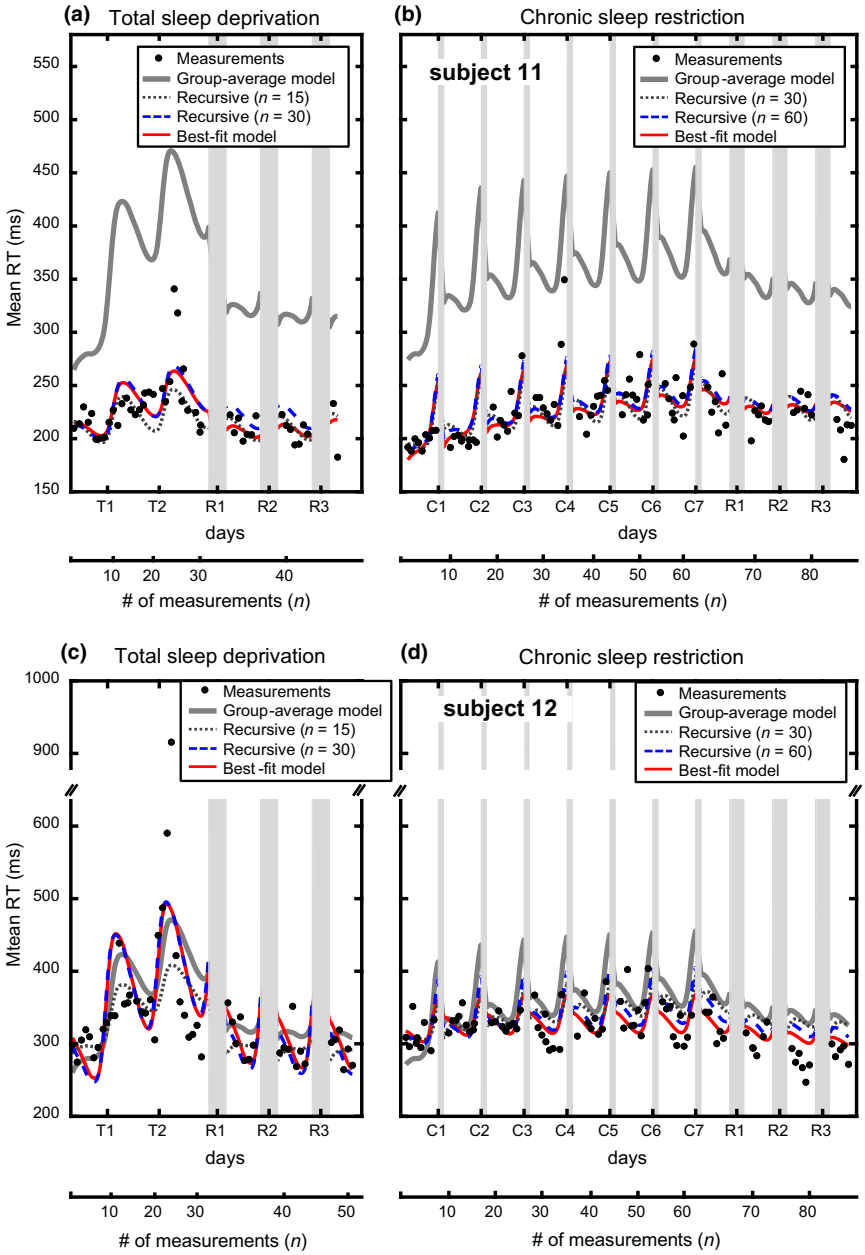

Figure 1. Snapshots of individualized recursive model predictions for two subjects with different responses to the sleep-loss challenges. Comparison of group-average model, best-fitted individualized model and individualized recursive learning under a 64-h total sleep deprivation challenge [T1–T2; (a) for subject 11 and (c) for subject 12] and a chronic sleep restriction challenge consisting of seven consecutive nights of 3 h of nightly time in bed (TIB) [C1–C7; (b) for subject 11 and (d) for subject 12]. Both challenges were followed by three recovery nights of 8 h of nightly TIB (R1–R3). Only the recursively learned models applied to the first 15 and 30 measurements under the total sleep deprivation challenge or 30 and 60 measurements under the chronic sleep restriction challenge are plotted. The grey-shaded vertical bars represent sleep episodes. RT: response time.

three parameters, U, κ and ϕ, each to exceed 0.80 (which indicates an almost perfect agreement, i.e. parameter convergence). For CSR, it required 63 measurements sampled every 2 h during wakefulness. We observed notable differences in the rate of learning of the different parameters. For both challenges, U reached almost perfect agreement considerably earlier (n = 12 for TSD and n = 42 for CSR) than did the two other parameters.

We also computed the relative RMSE between the recursive algorithm and the best-fitted model averaged across the 18 subjects (Fig. 3, bottom). For TSD, the recursive algorithm required 31 measurements sampled every 2 h to yield a relative RMSE of less than 10% (our arbitrary definition of UMP output convergence, which corresponds to ~4 ms for mean RT under TSD and CSR). For CSR, it required 44 measurements sampled every 2 h during wakefulness.

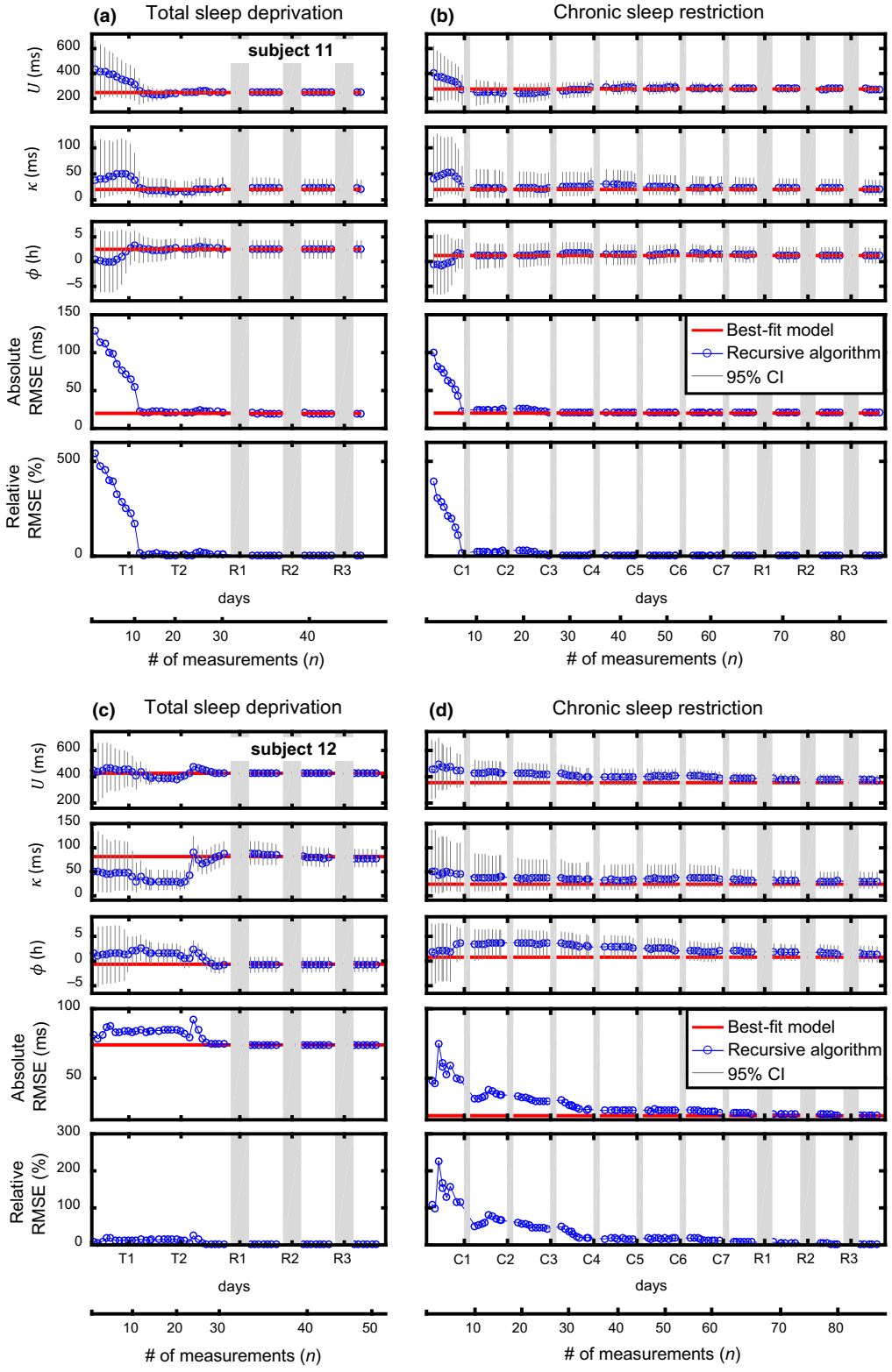

Figure 2. Temporal progression of model individualization for two subjects. Parameter estimates, their associated confidence intervals (CIs), the absolute and the relative root mean squared errors (RMSEs) as a function of psychomotor vigilance task measurements for a 64-h total sleep deprivation challenge [T1–T2; (a) for subject 11 and (c) for subject 12 and a chronic sleep restriction challenge consisting of seven consecutive nights of 3 h of nightly time in bed (TIB) [C1–C7; (b) for subject 11 and (d) for subject 12]. Both challenges were followed by three recovery nights of 8 h of nightly TIB (R1–R3). The red horizontal lines indicate the parameters and absolute RMSEs of the best-fitted model using all measurements. The grey-shaded vertical bars represent sleep episodes. U: upper asymptote of the homeostatic process; κ: circadian amplitude; ϕ: circadian phase.

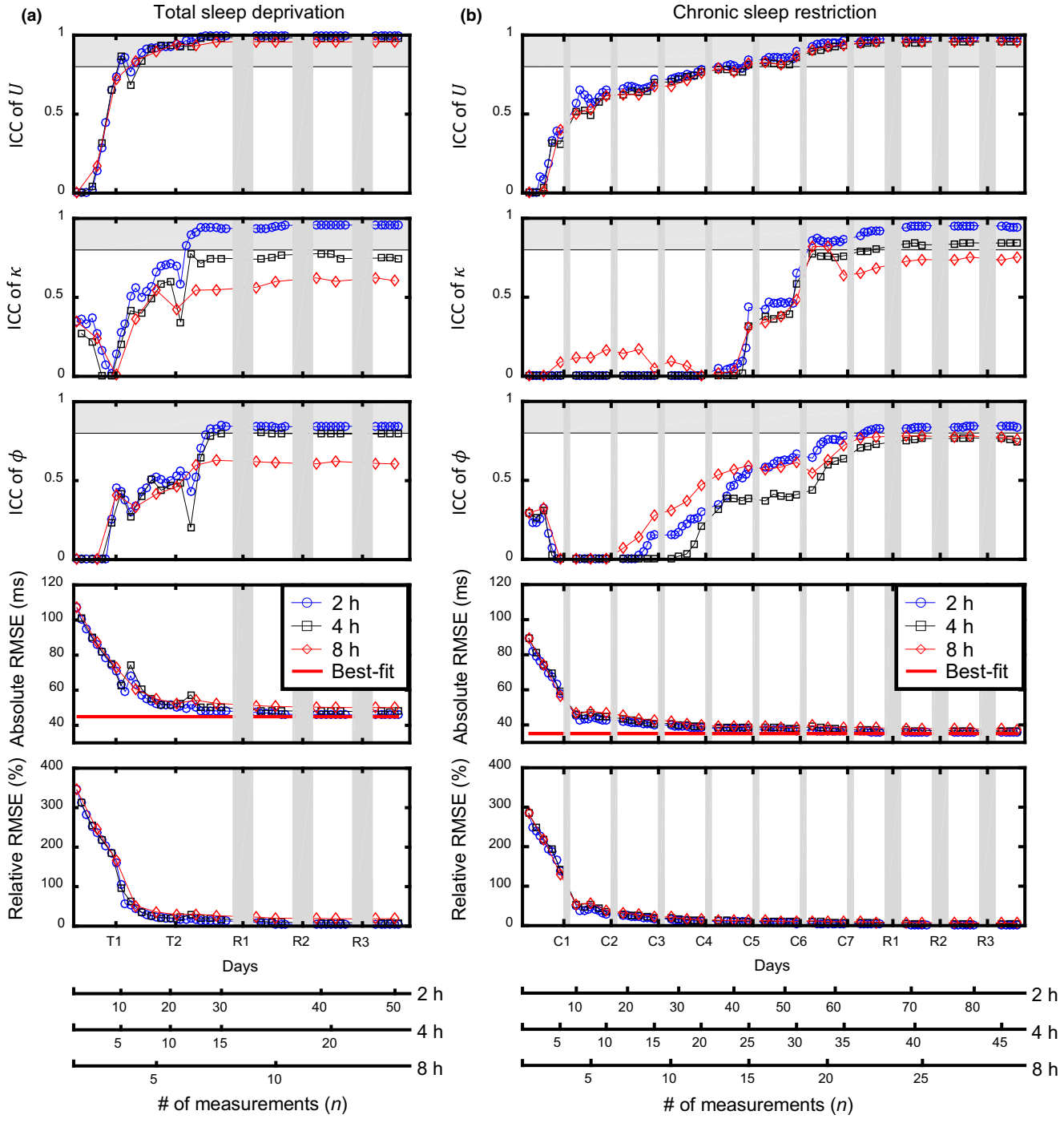

Figure 3. Convergence of model parameters across all 18 subjects. Convergence of parameter estimates and model predictions under a 64-h total sleep deprivation challenge (T1–T2; a) and a chronic sleep restriction challenge consisting of seven consecutive nights of 3 h of nightly time in bed (TIB) (C1–C7; b) as a function of PVT measurement frequency (every 2, 4 or 8 h). Both challenges were followed by three recovery nights of 8 h TIB (R1–R3). The red horizontal lines indicate the averaged absolute root mean squared error (RMSE) of the best-fitted models from the 18 subjects. The grey-shaded vertical bars represent sleep episodes. The grey-shaded horizontal bars in the plots for the intraclass correlation coefficients (ICCs) of parameters U, κ and ϕ represent the range of almost perfect (0.80–1.00) agreement, indicating that the parameter has converged. U: upper asymptote of the homeostatic process; κ: circadian amplitude; ϕ: circadian phase.

Impact of PVT measurement frequency on model individualization

To assess the impact of PVT measurement frequency on the rate of model individualization with the recursive algorithm, we down-sampled the data by a factor of two (to simulate 4-h sampling) or four (to simulate 8-h sampling) and repeated the analyses (Fig. 3). Reducing the frequency of PVT measurements from once every 2 h to once every 4 h did not affect the learning rate of the model parameters noticeably, except

Table 3 Comparison of five psychomotor vigilance task (PVT) statistics. Shown are the number of measurements required for the recursively learned parameters to converge and the models' output root mean squared errors (RMSEs) for total sleep deprivation (TSD) and chronic sleep restriction (CSR; entries within parentheses) averaged over the 18 subjects in our study

	Number of measurements to converge (ICC > 0.80)			*Absolute RMSE*	
PVT statistic	*U*	κ	ϕ	*Best-fitted model*	*Recursive algorithm**
Mean RT	12 (42)	22 (53)	27 (63)	45 (35) ms	46 (36) ms
Median RT	7 (26)	22 (69)	27 (78)	21 (23) ms	22 (23) ms
Slowest 10% RT	14 (34)	22 (62)	27 (77)	225 (160) ms	231 (164) ms
Speed	8 (23)	22 (56)	28 (–)	0.25 (0.30) s^{-1}	0.26 (0.31) s^{-1}
Lapses	4 (16)	22 (58)	26 (–)	3.91 (3.90) lapses	4.00 (3.97) lapses

ICC: intraclass correlation coefficient; RT: response time; *U*: upper asymptote of the homeostatic process; κ: circadian amplitude; ϕ: circadian phase.
*Using all available measurements: 51 for TSD and 85 for CSR.

for the ICC of the circadian amplitude κ under TSD and for the ICC of the circadian phase ϕ under CSR, which only reached levels of substantial agreement ($0.61 < \text{ICC} < 0.80$). However, down-sampling from 2 to 4 h did not noticeably affect the RMSEs. In contrast, reducing the measurement frequency further to once every 8 h had a more pronounced effect on the learning of the circadian rhythm parameters and, to a lesser extent, on the model outputs and RMSEs. These effects were more noticeable for TSD.

Analysis of additional PVT statistics

To determine whether the choice of PVT statistics affected our results, we repeated our analyses of the mean RT for four additional PVT statistics: (1) median RT, (2) slowest 10% RT, (3) speed and (4) lapses. Table 3 shows that the overall patterns discussed above, and repeated in the first row of the table for mean RT, are consistent with the results obtained for the other PVT statistics. The three UMP parameters (*U*, κ and ϕ) converged to an ICC of >0.80 within the available number of measurements for each of the PVT statistics, except for the circadian phase ϕ for speed and lapses under CSR, which only reached ICCs of 0.71 and 0.75, respectively, after 85 measurements. As in the case of mean RT, the recursive algorithm required fewer measurements for the UMP parameters to converge under the TSD challenge than it did under the CSR challenge. For TSD, all parameters converged within ~30 PVT measurements, whereas for CSR it took as many as 78 measurements. Using the complete set of measurements for each challenge, we obtained absolute RMSEs of the model outputs, averaged over the 18 subjects, which were indistinguishable between the recursively learned and best-fitted models for all statistics (Table 3).

Preservation of recursively learned parameters

We also confirmed that the recursively learned response to sleep loss for an individual under one sleep-loss challenge was preserved for another (the ICCs of the estimates of *U*, κ and ϕ recursively learned by using all data across the TSD and CSR challenges were 0.85, 0.53 and 0.48, respectively), consistent with a previous report (Ramakrishnan *et al.*, 2015). Follow-up analysis confirmed that, by using the parameter estimates of each individual at the end of one sleep-loss challenge as the initial priors for the same individual under another challenge, we could facilitate model individualization (Fig. 4). For example, the ICC of the most sensitive parameter *U* reached the region of almost perfect agreement (ICC >0.8) at the onset of learning immediately, and the RMSEs (both absolute and relative) during the initial phase of learning (first 10 PVT measurements) were considerably lower than those obtained with the original priors.

Recursive learning under various chronic sleep restriction conditions

To determine the extent to which the insights gained in the CSR analyses are generalizable to other challenges (especially less severe CSR schedules), we repeated our simulations for another CSR study, in which different subjects were challenged with seven consecutive nights of 3, 5, 7 or 9 h of nightly TIB (Belenky *et al.*, 2003). Although model individualization required considerably more time with 9-h TIB, the recursive algorithm progressively learned the responses to sleep loss with 3, 5 and 7 h of nightly TIB after ~1 week of four daily PVT measurements (Fig. 5).

DISCUSSION

Applications that use mobile computing platforms to predict an individual's neurobehavioural performance must be able to 'learn' the individual's trait-like response to sleep loss automatically over time. Here we proposed a recursive algorithm based on an extended Kalman filter to individualize in real time a validated model of performance. We used data from a cross-over design study of 18 subjects challenged with both TSD and CSR conditions to simulate real-time performance and assess the algorithm.

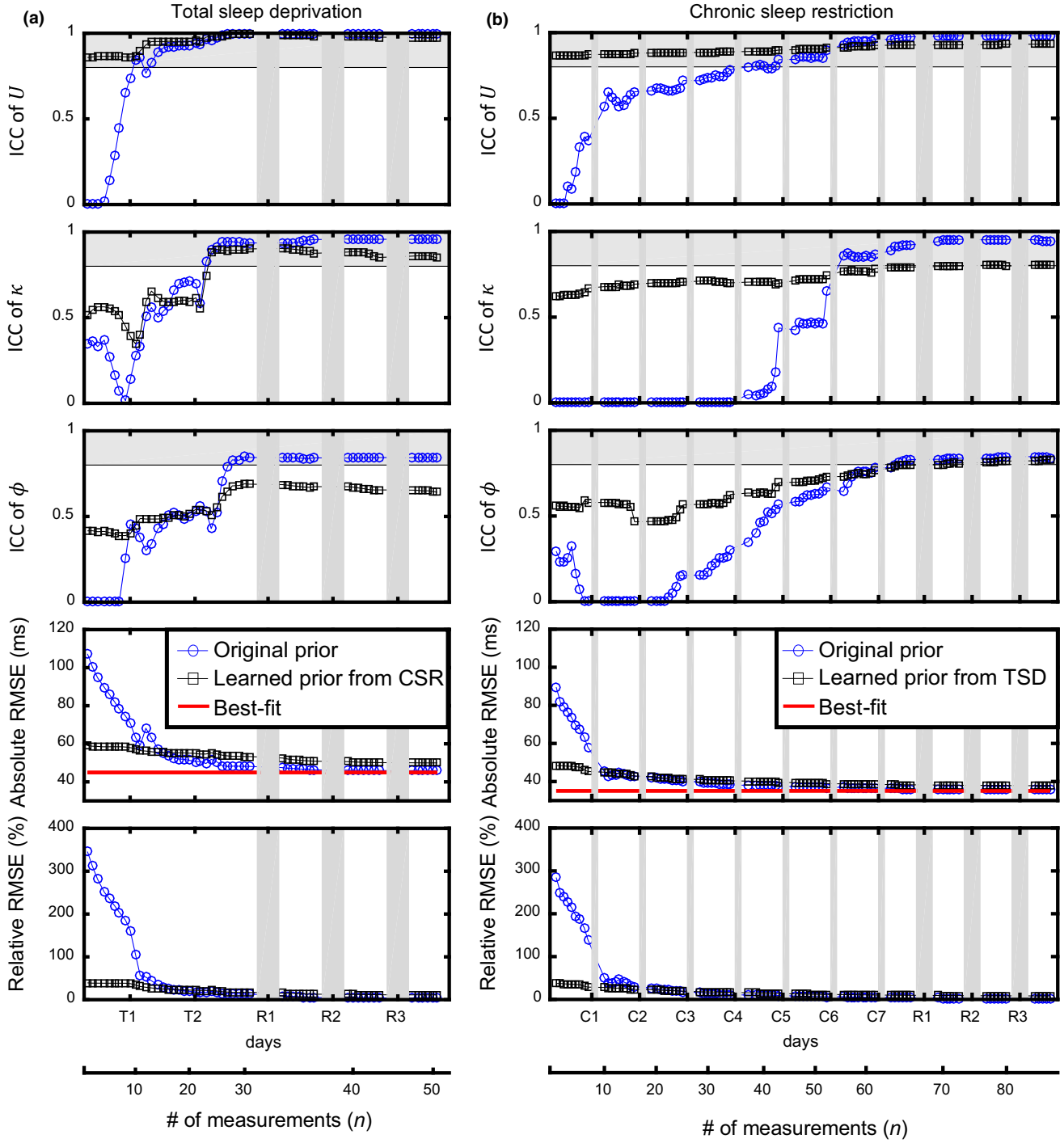

Figure 4. Continual learning across sleep-loss challenges. Comparison of the convergence of model individualization across all 18 subjects under a 64-h total sleep deprivation (TSD) challenge (a, T1–T2) and a chronic sleep restriction (CSR) challenge consisting of seven consecutive nights of 3 h of nightly time in bed (TIB) (b, C1–C7), using the original priors (as in Fig. 3) and the learned priors from one condition as the starting parameter values for the other condition. Both challenges were followed by three recovery nights of 8 h of nightly TIB (R1–R3). The grey-shaded vertical bars represent sleep episodes. The grey-shaded horizontal bars in the plots for the intraclass correlation coefficient for parameters U, κ and ϕ represent the range of almost perfect (0.80–1.00) agreement. ICC, intraclass correlation coefficient; RMSE: root mean squared error; U: upper asymptote of the homeostatic process; κ: circadian amplitude; ϕ: circadian phase.

We investigated three questions. First, we assessed how the model obtained by the recursive algorithm fared relative to that obtained by *post-hoc* fitting of the UMP with all PVT measurements. A comparison of the model parameters after using all PVT measurements to train the recursive algorithm with those obtained in the best-fitted model revealed small

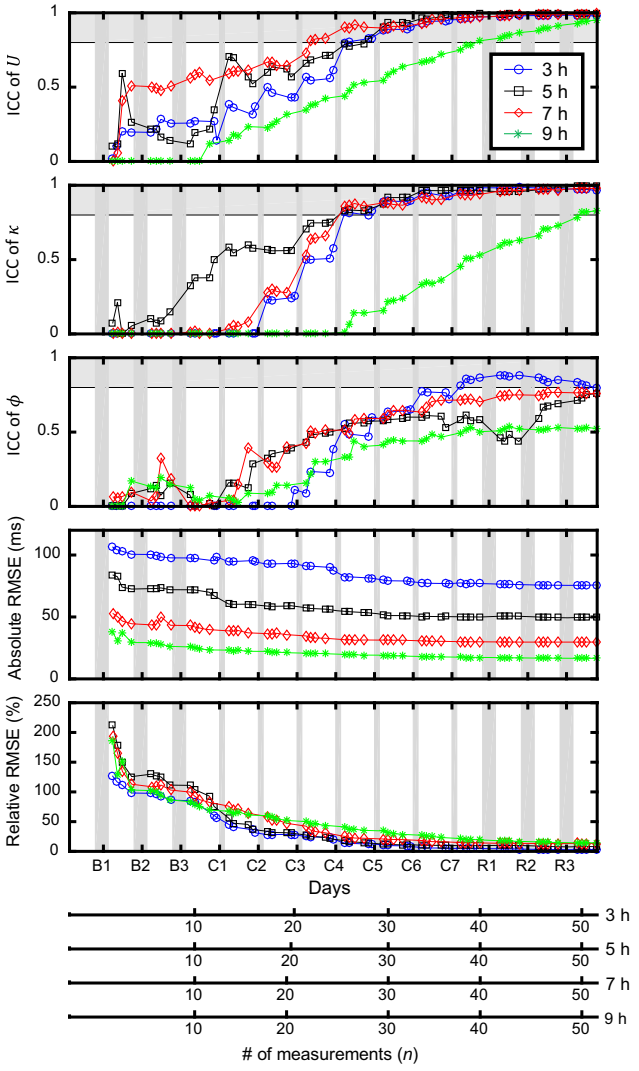

Figure 5. Convergence of model parameters under various chronic sleep restriction conditions. Convergence of parameter estimates and model predictions under various chronic sleep restriction conditions [seven nights of 3, 5, 7 or 9 h of nightly time in bed (TIB)]. Each challenge (C1–C7) was preceded (B1–B3) and followed (R1–R3) by three nights of 8 h of nightly TIB. The grey-shaded vertical bars represent sleep episodes (only the 3-h condition is shown during the sleep-loss challenges). The grey-shaded horizontal bars in the plots of the intraclass correlation coefficient for parameters U, κ and ϕ represent the range of almost perfect (0.80–1.00) agreement. ICC; intraclass correlation coefficient; RMSE: root mean squared error; U: upper asymptote of the homeostatic process; κ: circadian amplitude; ϕ: circadian phase.

differences (average absolute values of 14 ms for U, 6 ms for κ and 0.8 h for ϕ). In addition, a comparison of the model outputs obtained after using all measurements to train the recursive algorithm with those for the best-fitted model revealed negligible differences in absolute and relative RMSEs for five PVT statistics (Table 3). Overall, these results suggest that, at the end of a sleep-loss challenge, a model individualized by using one PVT measurement at a time converges ultimately to the same model as that produced by using all data at once in a *post-hoc* manner. We also confirmed that the recursively learned model parameters were preserved for a given individual across the two sleep-loss challenges (Fig. 4). Further validation by using a different data set (Fig. 5) suggested that the recursive algorithm could learn the individual's response to sleep loss progressively under less severe CSR conditions.

Secondly, we investigated the minimum number of PVT measurements needed to individualize the UMP under the TSD and CSR challenges. Our results showed unequivocally that this number was highly dependent upon the sleep-loss challenge. With the recursive algorithm, convergence of model parameters under 64 h of TSD required far fewer measurements than under 7 days of 3 h nightly TIB (Table 3). For TSD, each UMP parameter converged within 30 measurements sampled every 2 h for all PVT statistics (Table 3). In contrast, for CSR, convergence required ~70 measurements sampled every 2 h (with the exception of φ for the speed and lapses metrics, Table 3). There are two probable explanations for this difference. First, the pressure to sleep increases continuously during TSD but not during CSR. This results in faster performance degradation which, in turn, enables more rapid estimation of U (Fig. 3, top). Secondly, in contrast to CSR, in which collection of measurements is interrupted during sleep periods, the continual availability of measurements during TSD over entire 24-h periods makes it possible to estimate accurately κ and ϕ in fewer days (Fig. 3).

Finally, we investigated the impact of PVT measurement frequency on the ability of the recursive algorithm to learn an individual's trait-like response to sleep loss. Surprisingly, down-sampling of measurements from once every 2 h to once every 4 or 8 h had no apparent effect on the relative RMSEs for either challenge (Fig. 3, bottom). This is because down-sampling had almost no impact on the learning rate for U, which drives the UMP outputs (Fig. 3, top). However, as we reduced the sampling frequency, the learning rates for κ and φ plateaued, yielding suboptimal levels of agreement for TSD, particularly for 8-h sampling (Fig. 3). We did not observe such a plateau for CSR, presumably because the measurements available to estimate the circadian parameters accurately were insufficient under TSD (~15 for 4-h sampling and ~8 for 8-h sampling), and noticeably fewer than those under CSR (~38 for 4-h sampling and ~23 for 8-h sampling).

Importantly, the ability of the recursive algorithm to learn the model parameters depends not only upon the number of PVT measurements *per se* but also on the period during which the data are collected. For a fixed number of measurements, the parameter estimates are more accurate the greater the number of 24-h cycles included in the data collection period. For example, after 20 measurements in CSR, 8-h sampling achieved better levels of agreement for all three parameters than did 4-h sampling which, in turn, resulted in better levels of agreement than did 2-h sampling (Fig. 3).

A practical implication of these findings is that, to individualize the UMP, 4-h PVT sampling results in minimal decrements in model performance when compared with the more demanding 2-h sampling schedule, which is used commonly in laboratory studies. In addition, longer data collection periods improve the learning ability of the recursive algorithm.

One limitation of this study is that the results are based on a study of 18 healthy young adults. Hence, the extent to which the present findings are applicable to a heterogeneous, older population is unclear. Another limitation is that the results are based on PVT test statistics. Whether the findings can be extended to other neurocognitive measures of performance impairment remains to be investigated.

In summary, this work demonstrates the ability of a recursive algorithm to individualize UMP parameters in real time in a computationally efficient manner, allowing for the development of smartphone applications (apps) that can be customized to predict individual neurobehavioural performance. Our group is currently developing such an app, in which an individual's sleep/wake history, caffeine consumption and intermittent PVT measurements are used to continually update and individualize the UMP parameters, and thereby optimize predictions of the individual's future performance.

ACKNOWLEDGEMENTS

This work was sponsored by the Military Operational Medicine Research Area Directorate of the US Army Medical Research and Materiel Command, Fort Detrick, MD and by the US Department of Defense Medical Research and Development Program (Grant DMRDP_13200).

AUTHOR CONTRIBUTIONS

SR, JR and TJB conceived the research. JL, SR and SL developed the algorithms and performed the computations. TJB provided data for modelling and simulation. JL, SR, SL and JR wrote the paper.

CONFLICT OF INTEREST

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. This paper has been approved for public release with unlimited distribution.

REFERENCES

Arulampalam, M. S., Maskell, S., Gordon, N. and Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, 2002, 50: 174–188.

Belenky, G., Wesensten, N. J., Thorne, D. R. *et al.* Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose–response study. *J. Sleep Res.*, 2003, 12: 1–12.

Borbély, A. A. and Achermann, P. Sleep homeostasis and models of sleep regulation. *J. Biol. Rhythms*, 1999, 14: 557–568.

Chen, Z. Bayesian filtering: from Kalman filters to particle filters, and beyond. *Statistics*, 2003, 182: 1–69.

Dawson, D., Ian Noy, Y., Härmä, M., Åkerstedt, T. and Belenky, G. Modelling fatigue and the use of fatigue models in work settings. *Accid. Anal. Prev.*, 2011, 43: 549–564.

Hursh, S. R., Redmond, D. P., Johnson, M. L. *et al.* Fatigue models for applied research in warfighting. *Aviat. Space Environ. Med.*, 2004, 75: A44–A53.

Landis, J. R. and Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics*, 1977, 33: 159–174.

Mallis, M. M., Mejdal, S., Nguyen, T. T. and Dinges, D. F. Summary of the key features of seven biomathematical models of human fatigue and performance. *Aviat. Space Environ. Med.*, 2004, 75: A4–A14.

Olofsen, E., Dinges, D. F. and Van Dongen, H. P. Nonlinear mixed-effects modeling: individualization and prediction. *Aviat. Space Environ. Med.*, 2004, 75: A134–A140.

Rajaraman, S., Gribok, A. V., Wesensten, N. J., Balkin, T. J. and Reifman, J. Individualized performance prediction of sleep-deprived individuals with the two-process model. *J. Appl. Physiol.*, 2008, 104: 459–468.

Rajaraman, S., Gribok, A. V., Wesensten, N. J., Balkin, T. J. and Reifman, J. An improved methodology for individualized performance prediction of sleep-deprived individuals with the two-process model. *Sleep*, 2009, 32: 1377–1392.

Rajdev, P., Thorsley, D., Rajaraman, S. *et al.* A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation. *J. Theor. Biol.*, 2013, 331: 66–77.

Ramakrishnan, S., Lu, W., Laxminarayan, S. *et al.* Can a mathematical model predict an individual's trait-like response to both total and partial sleep loss? *J. Sleep Res.*, 2015, 24: 262–269.

Ramakrishnan, S., Wesensten, N. J., Balkin, T. J. and Reifman, J. A unified model of performance: validation of its predictions across different sleep/wake schedules. *Sleep*, 2016, 39: 249–262.

Rupp, T. L., Wesensten, N. J. and Balkin, T. J. Trait-like vulnerability to total and partial sleep loss. *Sleep*, 2012, 35: 1163–1172.

Seber, G. A. F. and Wild, C. J. *Nonlinear Regression*, 1st edn. Wiley-Interscience, Hoboken, NJ, 2003.

So, C. J., Quartana, P. J. and Ratcliffe, R. H. Caffeine efficacy across a simulated 5-day work week with sleep restriction. *Sleep*, 2016, 39(Abstract Suppl): A92.

Van Dongen, H. P., Baynard, M. D., Maislin, G. and Dinges, D. F. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep*, 2004, 27: 423–433.

Van Dongen, H. P., Mott, C. G., Huang, J. K., Mollicone, D. J., Mckenzie, F. D. and Dinges, D. F. Optimization of biomathematical model predictions for cognitive performance impairment in individuals: accounting for unknown traits and uncertain states in homeostatic and circadian processes. *Sleep*, 2007, 30: 1129–1143.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. Snapshots of parameter estimates for all 18 subjects.

Figure S1. Snapshots of the parameter estimates during recursive learning for all 18 subjects after $n = 15$, 30, and 51 measurements under total sleep deprivation (TSD, *A*) and after $n = 30$, 60, and 85 measurements chronic sleep restriction (CSR, *B*).

Data S2. Sensitivity analysis.

Table S1. The average absolute partial rank correlation coefficient (|PRCC|) and the average first-order effect from the extended Fourier amplitude sensitivity test over the entire sleep-loss challenge for the 8 model parameters, under a 64-h total sleep deprivation (TSD) challenge and a chronic sleep restriction (CSR) challenge consisting of seven consecutive nights of 3 h of nightly time in bed. Both challenges were followed by three recovery nights of 8 h of nightly time in bed.