

Individualized Performance Prediction During Total Sleep Deprivation: Accounting for Trait Vulnerability to Sleep Loss

Sridhar Ramakrishnan, Srinivas Laxminarayan, David Thorsley, Nancy J. Wesensten, Thomas J. Balkin, and Jaques Reifman

Abstract—Individual differences in vulnerability to sleep loss can be considerable, and thus, recent efforts have focused on developing individualized models for predicting the effects of sleep loss on performance. Individualized models constructed using a Bayesian formulation, which combines an individual's available performance data with *a priori* performance predictions from a group-average model, typically need at least 40 h of individual data before showing significant improvement over the group-average model predictions. Here, we improve upon the basic Bayesian formulation for developing individualized models by observing that individuals may be classified into three sleep-loss phenotypes: resilient, average, and vulnerable. For each phenotype, we developed a *phenotype-specific* group-average model and used these models to identify each individual's phenotype. We then used the phenotype-specific models within the Bayesian formulation to make individualized predictions. Results on psychomotor vigilance test data from 48 individuals indicated that, on average, ~85% of individual phenotypes were accurately identified within 30 h of wakefulness. The percentage improvement of the proposed approach in 10-h-ahead predictions was 16% for resilient subjects and 6% for vulnerable subjects. The trade-off for these improvements was a slight decrease in prediction accuracy for average subjects.

I. INTRODUCTION

NUMEROUS studies have demonstrated that there is significant inter-individual variability in psychomotor performance when humans are sleep deprived [1, 2]. In particular, it is believed that individuals can be broadly categorized into three sleep-loss phenotypes, *resilient*, *average*, and *vulnerable*, where the percentage of individuals in each category can vary from 20% to 40% in a given study

This work was supported, in part, by the Military Operational Medicine Research Area Directorate of the U.S. Army Medical Research and Materiel Command (USAMRMC), Ft. Detrick, MD.

S. Ramakrishnan is with the DoD Biotechnology High Performance Computing Software Applications Institute (BHS AI), Telemedicine and Advanced Medical Technology Research Center (TATRC), USAMRMC, Frederick, MD 21702 USA.

S. Laxminarayan is with the BHS AI, TATRC, USAMRMC, Frederick, MD 21702 USA.

D. Thorsley is with the BHS AI, TATRC, USAMRMC, Frederick, MD 21702 USA.

N. J. Wesensten is with the Center for Military Psychiatry and Neuroscience, Dept. of Behavioral Biology, Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD 20910 USA.

T. J. Balkin is with the Center for Military Psychiatry and Neuroscience, Dept. of Behavioral Biology, WRAIR, Silver Spring, MD 20910 USA.

J. Reifman is a Senior Research Scientist and Director of the BHS AI, TATRC, USAMRMC, ATTN: MCMR-TT, 504 Scott Street, Fort Detrick, MD 21702 USA (corresponding author; phone: 301-619-7915; fax: 301-619-1983; e-mail: jaques.reifman@us.army.mil).

due to the relatively small sample size of each investigation [3]. This realization has led to a shift in the development of biomathematical performance models away from “group-average” models, which represent the average performance of a group of individuals, and toward “individualized” models, where models are customized to capture the sleep-loss phenotype variability of each individual.

Recently, our group developed individualized performance models that use previous measurements of performance from a specific individual to customize the model, i.e., to adjust the model parameters, to that individual and estimate future performance values for a given prediction horizon [4, 5]. The model requires a minimum of 13 prior performance observations [4], but because we wish to individualize the model and make predictions as soon as the first measurement of performance becomes available, we employed a Bayesian approach [5]. Our Bayesian formulation combines performance data from the available measurements with *a priori* performance data, which, in their absence, are estimated from a *fixed* group-average prediction model.

Although our individualized performance models improved predictions by as much as 43% for a 10-h-ahead prediction horizon, the results indicate that the rate at which the model “learns” the sleep-loss phenotype of an individual is highly dependent on how representative the *a priori* group-average model is to the individual's phenotype [5]. In this study, we investigate whether in our Bayesian formulation we can accelerate the learning rate of the individualized models by using phenotype-specific (resilient, average, and vulnerable) *a priori* group-average models instead of a fixed, *a priori* group-average model. We used laboratory study data from 48 subjects exposed to 64.5 h of total sleep deprivation, where performance was measured through psychomotor vigilance tests (PVTs) [6].

II. METHODS

A. Individualized Two-Process Prediction Model

We used the two-process model of sleep-regulation [7] as the underlying model for our individualized prediction of performance impairment $P(k)$ due to sleep loss, where k is a discrete-time index. The model represents performance as an additive interaction of two processes [8]: the sleep homeostatic Process S , which increases exponentially with time awake and decreases exponentially with time asleep [9], and the circadian Process C , which is independent of sleep/wake history [10]. During total sleep deprivation, performance $P(k)$ is described as follows [4]:

$$P(k) = \alpha - \alpha S_0 \exp[-(k-1)\rho T_s] + \beta \sum_{i=1}^5 a_i \sin\left\{\frac{2\pi}{\tau} i[(k-1)T_s + \phi]\right\}, \quad (1)$$

where α and β are parameters that control the relative effect of the Processes S and C on performance, respectively, ρ denotes the buildup rate of homeostatic pressure, T_s denotes the sampling period, S_0 denotes the initial homeostatic state, and ϕ denotes the initial circadian phase. These five parameters are individual specific, constant, and have unknown values. The parameter τ denotes the time period of the circadian clock (24 h), and the parameters a_i , $i = 1, \dots, 5$, define the amplitudes of the five harmonics of Process C ($a_1 = 0.97$, $a_2 = 0.22$, $a_3 = 0.07$, $a_4 = 0.03$, and $a_5 = 0.001$) [10].

In our prior work [5], we used a Bayesian formulation, which combines available performance data from an individual with his/her *a priori* performance predictions from a group-average model, to obtain individualized two-process model parameter estimates. The process of model individualization begins when the first performance measurement is taken and an individual's parameter estimates are updated recursively as each new performance measurement becomes available. After any number of measurements, we used the individualized two-process model obtained by this procedure to predict $P(k)$ for any desired future value of k .

In our prior work, the group-average model, also based on Eq. (1), has *fixed* model parameters obtained from Van Dongen et al. [2]. In this paper, however, we used *phenotype-specific* group-average model parameters instead of *fixed* parameters to accelerate the rate at which models learn an individual's phenotype, and thus, improve prediction accuracy.

B. Phenotype-specific Group-average Model

We used performance data from 48 individuals [11] to formulate three phenotype-specific group-average models corresponding to three sleep-loss phenotypes. To do so, we first used the K -means clustering scheme [12], a popular unsupervised learning algorithm, to classify the temporal performance profiles of the 48 subjects into three classes, which we then labeled as *resilient*, *average*, and *vulnerable* based on the energy of their centroids. For each class, we randomly separated the individuals into training (~60%) and validation (~40%) sets. Using data from each of the three training sets, we developed a phenotype-specific group-average model for each class by performing mixed-effects regression [2, 13]. Using this procedure, we obtained the means and variances of the group-average model parameters for each of the three phenotype-specific classes.

We assumed that individuals in the validation sets had unknown sleep-loss phenotypes that needed to be identified from their performance measurements. To use an individual's phenotype-specific group-average model parameters, we determined his/her phenotype by computing the log-likelihood distance $d_p(n)$ between the available measurements and the three phenotype-specific group-average model predictions as follows:

$$d_p(n) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log|\Sigma_{n,p}| + \frac{1}{2} \left[(y_n - \mu_{n,p})^T \Sigma_{n,p}^{-1} (y_n - \mu_{n,p}) \right], \quad (2)$$

where n denotes the number of available measurements, y_n denotes the vector of available performance measurements, $\mu_{n,p}$ and $\Sigma_{n,p}$ denote the mean vector and covariance matrix of the n predictions of the p -th phenotype-specific group-average model, respectively, and p is an index corresponding to the resilient, average, or vulnerable phenotype classes. We then classified an individual as a member of the phenotype corresponding to the smallest d_p .

Each time a new performance measurement became available, we repeated the above scheme of identifying the unknown phenotype and using the corresponding phenotype-specific group-average model predictions to individualize the model. If an individual is close to the boundary between two classes, his/her phenotype classification may change as new measurements become available. If an individual's phenotype was reclassified, we switched the phenotype-specific group-average model accordingly to make predictions.

C. Model Prediction Performance

We compared the performance predictions for a desired horizon obtained using the proposed scheme with those obtained using the previous individualization approach that uses a fixed group-average model. To obtain the fixed model in this paper, we used the same training sets as described in Section II.B, pooled in data from all three classes, and performed mixed-effects regression to obtain the fixed group-average model parameters.

We used the root mean squared error (RMSE) between the data and the model predictions to compare the prediction accuracy of the two approaches. We performed 20 cross-validation trials, with each trial using a randomly selected set of 28 individuals for training and the remaining 20 individuals for validation. We then computed an average RMSE estimate of the prediction errors for each of the phenotype classes across the 20 validation sets.

III. RESULTS

A. Study Data

We validated the phenotype-specific prediction methodology using data from a controlled laboratory experiment [11]. In this experiment, 48 healthy adult subjects were kept awake for 64.5 h before being administered various pharmacological countermeasures to fatigue. Here, we considered only the data collected before the countermeasures were applied. Each subject completed a 10-min PVT session every 2 h, starting one hour after waking at 07:00 on the first day and finishing at 00:00 on the fourth day. During the 64.5-h time period of continuous wakefulness, a total of 32 PVT sessions were administered to each subject. The study was approved by the Walter Reed Army Institute of Research Human Use Committee and the

United States Army Medical Research and Materiel Command Human Subjects Review Board. Written informed consent was obtained from all subjects prior to their participation.

We used PVT lapses, defined by the number of response times greater than 500 ms, as our metric to quantify performance impairment. A larger number of lapses during a PVT session indicates greater impairment.

B. Phenotype Identification Accuracy

Figure 1 shows the results of the fixed model generated using all 28 subjects in the training set and the three phenotype-specific models generated using only the resilient, average, and vulnerable subjects. The resilient individual model predicts fewer PVT lapses than the fixed model, the vulnerable individual model predicts more PVT lapses than the fixed model, and the average individual model predicts about the same number of PVT lapses as the fixed model.

The performance of the phenotype-specific individualized model relies on accurate phenotype identification, especially when few measurements are available. We assessed the accuracy of phenotype identification as a function of the number of available measurements. Figure 2 shows the fraction of subjects in the validation set, averaged across 20 cross-validation trials, which were correctly classified.

We observe from Fig. 2 that the accuracy of phenotype identification is ~70% after 10 h of wakefulness and rapidly improves to ~85% after 30 h. Also, by 40 h, or equivalently with 20 available measurements, the phenotype detection accuracy is ~90%. The largest increase in accuracy, from the 8th to the 11th measurement, correspond to the time period (15-21 h) at which the phenotype-specific models were most separated from each other due to the circadian variation in performance (cf. Fig. 1).

C. Individualized Model Predictions

Table I compares the 6-, 10-, and 24-h-ahead prediction accuracy of the original approach, which uses a fixed group-average model, and the proposed approach, which uses a phenotype-specific group-average model. The two

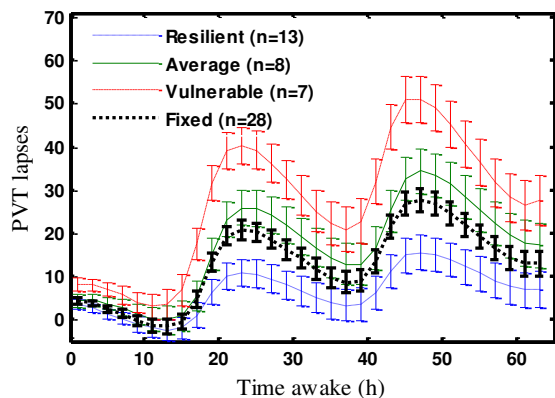


Figure 1. Phenotype-specific and fixed group-average models (mean \pm standard error) obtained from a training set of psychomotor vigilance test (PVT) lapse data from 28 subjects. The standard errors were obtained from the diagonal elements of the models' covariance matrices.

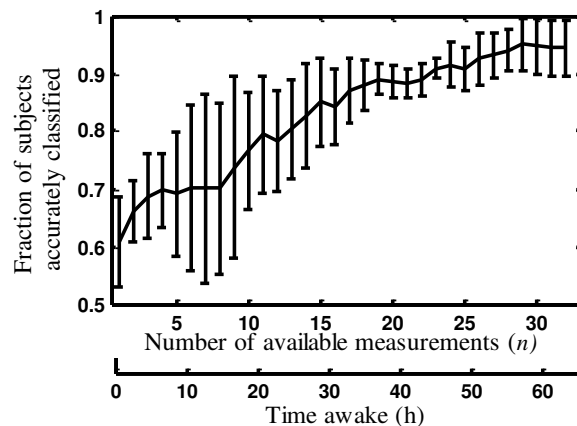


Figure 2. Accuracy of phenotype identification (mean \pm standard deviation) obtained by averaging over 20 cross-validation trials.

approaches are compared for each of the three sleep-loss phenotypes in terms of the RMSE (in PVT lapses) between the data and model predictions. For each of the prediction horizons, the phenotype-specific models performed better than the fixed model for the resilient subjects; for the 10- and 24-h prediction horizons, the improvement in performance was 16% and 20%, respectively, and reached statistical significance ($P < 0.05$) using a paired, two-sided sign test [14]. For the vulnerable subjects, the phenotype-specific model outperformed the fixed model, but the improvements of 6% and 8% for the 10- and 24-h prediction horizons, respectively, did not reach statistical significance. Conversely, for the average subjects, the phenotype-specific model performed slightly worse than the fixed model, but the difference again did not reach statistical significance.

Figure 3 shows the measured PVT performance (in PVT lapses), the fixed model 10-h-ahead prediction, and the phenotype-specific model 10-h-ahead prediction for an individual from each of the three phenotypes. We observe that the decrease in RMSE of ~2.5 lapses for resilient and vulnerable subjects using the phenotype-specific model is largely due to the improvement in predictions when the subjects were awake from 18 to 28 h. For the average subject, however, we observe no significant ($P \geq 0.05$) difference between the two model's predictions.

TABLE I. AVERAGE ROOT MEAN SQUARED ERRORS (RMSES) IN 6-, 10-, AND 24-H-AHEAD PERFORMANCE (PSYCHOMOTOR VIGILANCE TEST LAPSES) PREDICTIONS USING FIXED AND PHENOTYPE-SPECIFIC MODELS. NUMBERS IN PARANTHESES ARE STANDARD ERRORS IN THE RMSES. (* INDICATES $P < 0.05$ BASED ON A PAIRED, TWO-SIDED SIGN TEST).

Subject Phenotype	Fixed model			Phenotype-specific model		
	6 h	10 h	24 h	6 h	10 h	24 h
Resilient (n=22)	7.54 (0.36)	8.27 (0.28)	9.53 (0.32)	6.67 (0.55)	6.96* (0.51)	7.64* (0.55)
Average (n=14)	11.01 (0.78)	11.49 (0.77)	12.91 (1.13)	11.68 (0.78)	12.47 (0.79)	14.50 (1.06)
Vulnerable (n=12)	15.40 (0.85)	16.86 (1.04)	18.89 (1.27)	14.87 (0.74)	15.93 (0.85)	17.29 (1.24)

DISCUSSIONS AND CONCLUSION

In this work, we developed a set of *phenotype-specific* group average models based on a previously observed set of individual sleep-loss phenotypes. Using these *phenotype-specific* models, we improved upon the traditional Bayesian approach of using a single *fixed* model as a prior estimate for all individuals. Our proposed approach showed significant ($P < 0.05$) improvement in the accuracy of predictions for resilient subjects, modest improvement in accuracy for vulnerable subjects, and a slight decrease in accuracy for average subjects. We believe that the slight decrease in accuracy, due to the similarity between the *fixed* model and the *average* phenotype model, is a small trade-off in return for the much-improved predictions for resilient individuals.

For our proposed approach to perform most effectively, it is essential that an individual's phenotype be accurately determined as early as possible. Here, we use the minimum log-likelihood distance as our metric for determining an individual's phenotype. Alternate methods for determining the phenotype, such as the sequential probability ratio test [15], may provide rapid detection of phenotype and we intend to investigate their effectiveness as well.

Another avenue for future investigation is the effect of metric choice on phenotype selection. PVT lapses are just one of many metrics available for evaluating PVT performance [3]. Alternative metrics may show additional separation between the phenotypes after few measurements, resulting in more rapid phenotype classification and more

accurate performance predictions. We intend to investigate the properties of different PVT metrics in order to optimize the efficacy of our approach.

DISCLAIMER

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense.

REFERENCES

- [1] R. Leproult, E.F. Colecchia, A.M. Berardi, R. Stickgold, S.M. Kosslyn, and E. Van Cauter, "Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated," *Am J Physiol Regul Integr Comp Physiol*, vol. 284, no. 2, pp. R280-290, 2003.
- [2] H.P. Van Dongen, C.G. Mott, J.K. Huang, D.J. Mollicone, F.D. McKenzie, and D.F. Dinges, "Optimization of biomathematical model predictions for cognitive performance impairment in individuals: accounting for unknown traits and uncertain states in homeostatic and circadian processes," *Sleep*, vol. 30, no. 9, pp. 1129-1143, 2007.
- [3] S. Rajaraman, S. Ramakrishnan, D. Thorsley, N.J. Wessenden, T.J. Balkin, and J. Reifman, "A new metric for quantifying performance impairment on the psychomotor vigilance test," *J Sleep Res*, in press, 2012.
- [4] S. Rajaraman, A.V. Gribok, N.J. Wessenden, T.J. Balkin, and J. Reifman, "Individualized performance prediction of sleep-deprived individuals with the two-process model," *J Appl Physiol*, vol. 104, no. 2, pp. 459-468, 2008.
- [5] S. Rajaraman, A.V. Gribok, N.J. Wessenden, T.J. Balkin, and J. Reifman, "An improved methodology for individualized performance prediction of sleep-deprived individuals with the two-process model," *Sleep*, vol. 32, no. 10, pp. 1377-1392, 2009.
- [6] D. Dinges and J. Powell, "Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations," *Behavior Research Methods*, vol. 17, no. 6, pp. 652-655, 1985.
- [7] A.A. Borbely, "A two process model of sleep regulation," *Hum Neurobiol*, vol. 1, no. 3, pp. 195-204, 1982.
- [8] P. Achermann and A.A. Borbely, "Simulation of daytime vigilance by the additive interaction of a homeostatic and a circadian process," *Biol Cybern*, vol. 71, no. 2, pp. 115-121, 1994.
- [9] S. Daan, D.G. Beersma, and A.A. Borbely, "Timing of human sleep: recovery process gated by a circadian pacemaker," *Am J Physiol*, vol. 246, no. 2 Pt 2, pp. R161-183, 1984.
- [10] P. Achermann and A.A. Borbely, "Combining different models of sleep regulation," *J Sleep Res*, vol. 1, no. 2, pp. 144-147, 1992.
- [11] N.J. Wessenden, W.D. Killgore, and T.J. Balkin, "Performance and alertness effects of caffeine, dextroamphetamine, and modafinil during sleep deprivation," *J Sleep Res*, vol. 14, no. 3, pp. 255-266, 2005.
- [12] T. Hastie, R. Tibshirani, and J.H. Friedman, *The elements of statistical learning data mining, inference, and prediction*, New York, NY: Springer, 2009.
- [13] E. Olofsen, D.F. Dinges, and H.P. Van Dongen, "Nonlinear mixed-effects modeling: individualization and prediction," *Aviat Space Environ Med*, vol. 75, no. 3 Suppl, pp. A134-140, 2004.
- [14] J.H. Zar, *Biostatistical analysis*, Upper Saddle River, N.J.: Prentice Hall, 1999.
- [15] L. Chen, A.T. Reisner, X. Chen, A.V. Gribok, and J. Reifman, "Standard diagnostic test characteristics are insufficient in the assessment of continuous patient monitoring," *Medical Decision Making*, in press, 2012.

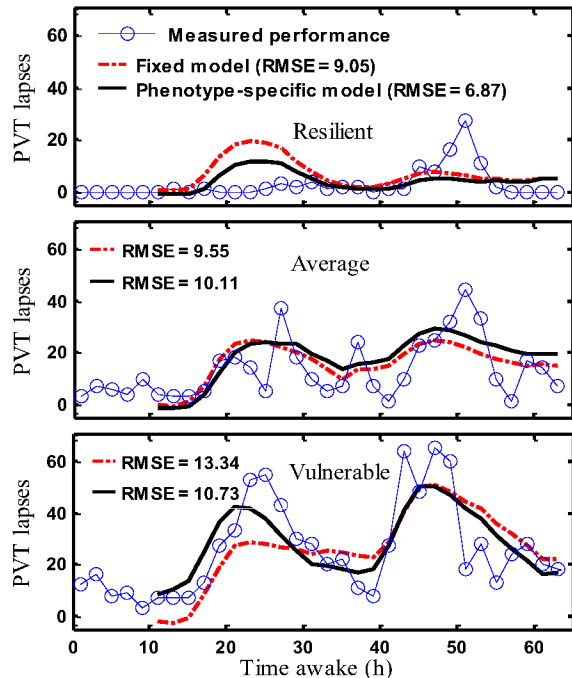


Figure 3. Comparison of 10-h-ahead psychomotor vigilance test (PVT) lapse predictions of individualized models that used the *fixed* and *phenotype-specific* group-average models for subjects of three sleep-loss phenotypes: *resilient* (top), *average* (middle), and *vulnerable* (bottom). Root mean squared errors (RMSEs) between the data and the model predictions are also provided.