

Moving Towards Individualized Performance Models

Commentary on Van Dongen HPA; Mott CG; Huang JK; Mollicone DJ; McKenzie FD; Dinges DF. 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(9):1129-1143.

Jaques Reifman, PhD; Srinivasan Rajaraman, PhD; Andrei V. Gribok, PhD

Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD

ONE PROMISING APPROACH TO BETTER MANAGE THE EFFECTS OF SLEEP DEPRIVATION ON PERFORMANCE IS THE USE OF BIOMATHEMATICAL MODELING TOOLS. However, owing to large inter-individual performance variability in humans exposed to similar sleep restrictions, models developed to date to predict group-average behavior have limited operational applicability. In this month's issue of *SLEEP*, Van Dongen et al. address this pressing issue by proposing a modeling approach for predicting fatigue and performance at an *individual-specific* level for humans continuously deprived of sleep.¹ Their approach is applicable when initial conditions, i.e., initial homeostat and circadian phase, are uncertain, and addresses another important performance modeling limitation² in that it attempts to quantify the model's prediction accuracy by estimating prediction error bounds in the form of "confidence intervals."

Together with previous work by this group,³ the paper by Van Dongen et al.¹ offers a refreshing departure from existing biomathematical models of performance. Individual-specific prediction is achieved by continually adapting the 5 parameters of the well-known two-process model of sleep regulation,^{4,5} so that, over time, its parameters are tuned to the individual being modeled. To tune a model to an individual, it is necessary to measure some physiologic biomarker from the individual and use it as part of the inputs to the model. In the absence of known predictive physiologic biomarkers of performance,⁶ previous performance measures of the individual being modeled, in the form of psychomotor vigilance tests (PVT), are fed back to tune the model and predict that individual's future performance. As noted in their paper, for practical implementation, the performance measures need to be automatically and passively obtained because in most operational environments it is not practical to interrupt a given activity to perform a PVT.

The Van Dongen et al. modeling approach is comprised of two steps.¹ In the first step, given performance measurements from

a group of individuals over the duration of a total sleep deprivation study, the two-process model of sleep regulation is cast in a mixed-effects regression framework that allows the decoupling of inter- and intra-individual variability of the performance measurements. This procedure results in probability distributions for the parameters of the two-process model and their associated group-average values, which are used as the starting point for predictions of unstudied individuals. While the mixed-effects procedure is effective in separating out the sources of variability in the data and yielding probability distributions for the model parameters that account solely for the effects of inter-individual variability, it inherently assumes that the group-average data are representative of the unstudied individuals we wish to predict. This assumption has practical implications. It restricts the types of individuals to whom the values of the group-average parameters may be applicable; data collected from young, healthy individuals may not be predictive of older individuals. Also, the level of noise in the group-average data needs to be similar to that of the individuals we want to predict. Hence, field-collected performance data are needed to obtain group-average parameters to predict individuals in an operational environment. Violation of this assumption, within the context of the Bayesian inference method used in their approach, may lead either to slow convergence of the learning process or large variance of the parameter estimates.

In the second step, Bayesian inference is applied to tune the model to an individual. In this data-learning algorithm, prior information from the probability distribution of the group-average parameters is balanced against information obtained from measured performance data from the individual being modeled and represented by a likelihood function. As each new performance observation becomes available, the likelihood function is recomputed, the 5 parameters are adapted, and the model is used to predict (up to 24 hours ahead) performance impairment levels of 3 subjects involved in an 88-hour total sleep deprivation study.

Several aspects of the results are counterintuitive, and require further consideration. Learning-from-data algorithms possess convergence properties that are not distinctly observed in the paper.¹ The performance predictions should become increasingly more accurate as more individualized data become available, indicating that the algorithm is continually learning, and shorter-horizon predictions should be more accurate than longer ones. As illustrated in Figure 2, predictions for Subject A at 36 hours of wakefulness are more accurate when performed at 12 hours than when performed (8 hours later) at 20 hours. Like other similar results, this seems counterintuitive. However, due to data uncertainty and other considerations, it is difficult to ascertain the properties of their, or any other, algorithm without knowing the true value of the results the algorithm is attempting to predict.

Disclosure Statement

Drs. Reifman, Rajaraman, and Gribok have indicated no financial conflicts of interest. Dr. Reifman was the Army's technical monitor (Contracting Officer's Representative) for the grant leading to the development of the biomathematical model discussed in this editorial.

Submitted for publication May, 2007

Accepted for publication May, 2007

Address correspondence to: Jaques Reifman, PhD, Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, MCMR-ZB-T, 363 Miller Drive, Fort Detrick, MD 21702; Tel: (301) 619-7915; Fax: (301) 619-1983; E-mail: jaques.reifman@us.army.mil

Van Dongen et al.¹ also attempt to quantify the accuracy of the predictions by estimating “confidence intervals” about the predictions. This effort should be commended, as prediction error bounds provide a measure of reliability of point estimates without which we do not know the extent the predictions should be trusted. However, the way the estimated intervals are computed is not in line with Bayesian procedures, which require knowledge of the distribution of the predicted values in the form of a predictive density function.⁷ This may be the reason why the estimated intervals show counterintuitive behavior throughout the prediction timeline. Figure 2 shows that for predictions performed up to 20 hours of wakefulness (top three rows), the width of the confidence intervals is smaller at the end of the prediction horizons than at some earlier horizon. For example, the predictions for all 3 subjects performed at 12 hours of wakefulness (second row) show a consistent and considerably smaller confidence interval at 36 hours (24-hour prediction horizon) than at some earlier time, say, 24 hours (12-hour prediction horizon). This implies that prediction uncertainty decreases with increasing horizons, which is nonsensical. Also, at later prediction times (bottom two rows in Figure 2), the width of the intervals remain constant across subjects and, surprisingly, seem to be independent of the variability in the data.⁸ One would expect the uncertainty in the predictions for Subject B, who shows little performance variability throughout the study, to be significantly smaller than those for Subjects A and C, who possess very noisy data, but they are not.

Assuming that measured performance data come from the two-process model, one way to unambiguously evaluate the characteristics and convergence properties of the Van Dongen et al. algorithm is through simulated data. The performance of an individual could be simulated by running the two-process model with fixed (known) parameter values and superimposing selected levels of random noise. Applying the algorithm of Van Dongen et al. to these simulated data, we could characterize the behavior of the estimated prediction error bounds to different conditions and determine whether the model parameter estimates converge to the true values, the rate of their convergence, and the effects of data noise in the parameter estimates and performance predictions. Moreover, it would allow for the evaluation of the bias-variance trade-off of their algorithm by determining the prediction accuracy (bias) and precision (variability) of the parameter estimates.

What steps are needed to improve the development of operationally useful, individual-specific biomathematical models of performance for humans exposed to sleep restrictions? First, because measurements from an individual are needed to predict that individual, it is imperative to identify biomarkers predictive of performance impairment, in particular ones that can be noninvasively and passively measured under unstructured operational environments. Second, given recent advancements in computer animation, simulations of real-world tasks representing scenarios, such as navigating a helicopter in an unknown terrain or distinguishing between friends and foes in a battlefield situation, should be used as the platform from which to attain measures of performance instead of general-purpose reaction tests that may not be a good indicator of performance in real-world tasks. Third, new mathematical model formulations may be needed and such formulations should be checked to assure that they fully capture the physiologic phenomena they intend to predict. Finally, the developed models need to be thoroughly validated through simulated data to guarantee that the models possess expected convergence properties.

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.

ACKNOWLEDGMENTS

This work was funded, in part, by the Military Operational Medicine and the Combat Casualty Care Research Area Directorates of the U.S. Army Medical Research and Materiel Command, Ft. Detrick, Maryland.

REFERENCES

1. Van Dongen HPA, Mott CG, Huang JK, Mollicone DJ, McKenzie FD, Dinges DF. 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;30:1129-43.
2. Reifman J. Alternative methods for modeling fatigue and performance. *Aviat Space Environ Med* 2004;75:A173-80.
3. Olofsen E, Dinges DF, Van Dongen HPA. Nonlinear mixed-effects modeling: individualization and prediction. *Aviat Space Environ Med* 2004;75:A134-40.
4. Borbély AA. A two process model of sleep regulation. *Hum Neurobiol* 1982;1:195-204.
5. Daan S, Beersma, DGM, Borbély AA. Timing of human sleep: recovery process gated by a circadian pacemaker. *Am J Physiol Regul Integr Comp Physiol* 1984;15:R164-78.
6. Afari N, Buchwald D. Chronic fatigue syndrome: a review. *Am J Psychiatry* 2003;160:221-36.
7. Berger JO. *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag, 1985.
8. Oleg' NO, Gribok AV, Reifman J. Error bounds for data-driven models of dynamical systems. *Comput Biol Med* 2007;37:670-9.