# An Improved Methodology for Individualized Performance Prediction of Sleep-Deprived Individuals with the Two-Process Model

Srinivasan Rajaraman, PhD[1]; Andrei V. Gribok, PhD[1]; Nancy J. Wesensten, PhD[2]; Thomas J. Balkin, PhD[2]; Jaques Reifman, PhD[1]

[1]Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD; [2]Department of Behavioral Biology, Walter Reed Army Institute of Research, Silver Spring, MD

We present a method based on the two-process model of sleep regulation for developing individualized biomathematical models that predict performance impairment for individuals subjected to total sleep loss. This new method advances our previous work in two important ways. First, it enables model customization to start as soon as the first performance measurement from an individual becomes available. This was achieved by *optimally* combining the performance information obtained from the individual's performance measurements with *a priori* performance information using a Bayesian framework, while retaining the strategy of transforming the nonlinear optimization problem of finding the optimal estimates of the two-process model parameters into a series of linear optimization problems. Second, by taking advantage of the linear representation of the two-process model, this new method enables the analytical computation of statistically based measures of reliability for the model predictions in the form of prediction intervals. Two distinct data sets were used to evaluate the proposed method.

Results using simulated data with superimposed white Gaussian noise showed that the new method yielded 50% to 90% improvement in parameter-estimate accuracy over the previous method. Moreover, the accuracy of the analytically computed prediction intervals was validated through Monte Carlo simulations. Results for subjects representing three sleep-loss phenotypes who participated in a laboratory study (82 h of total sleep loss) indicated that the proposed method yielded individualized predictions that were up to 43% more accurate than group-average prediction models and, on average, 10% more accurate than individualized predictions based on our previous method.

**Keywords:** Total sleep deprivation, cognitive performance impairment, individualized modeling, parameter estimation, performance prediction

**Citation:** Rajaraman S; Gribok AV; Wesensten NJ; Balkin TJ; Reifman J. An improved methodology for individualized performance prediction of sleep-deprived individuals with the two-process model. *SLEEP* 2009;32(10):1377-1392.

---

HISTORICALLY, BIOMATHEMATICAL PERFORMANCE PREDICTION MODELS HAVE FOCUSED ON PREDICTING GROUP-AVERAGE PERFORMANCE IMPAIRMENT.[1-3] This modeling strategy necessarily de-emphasizes individual differences in resilience and vulnerability to sleep loss and contradicts recent findings that indicate significant and systematic differences among individuals.[4-7] For example, on a psychomotor vigilance test (PVT) scale,[8] restricted-sleep laboratory studies show that interindividual differences account for nearly 70% of the total variance in performance among a group of individuals, and that the differences are trait-like, stable, and innate to an individual.[5] Hence, even if a group-average model were capable of accurately predicting mean-group performance, such a model would be of limited benefit without knowing how this translates into predictions at an individual level.[9]

Recently, two methods have been proposed for developing individual-specific performance prediction models for individuals subjected to total sleep loss with uncertain initial states (i.e., initial homeostatic pressure to sleep and circadian phase).[10,11] Both methods employ the two-process model of sleep regulation[12,13] as the underlying parametric model structure and, because of the absence of quantitative individual biomarkers of performance,[14,15] use previously collected PVT data from an individual to customize the model for that individual. These methods, which involve the minimization of the error between the model fit and a set of performance measurements from a specific individual, nonetheless differ in the techniques used to adapt the two-process model parameters to an individual. The method proposed by Van Dongen et al.[10] is conceptually based on the earlier work of Olofsen et al.,[16] with the original exponential model substituted by the two-process model of sleep regulation. Using a mixed-effects regression framework, they separate out inter- and intra-individual variability in previously recorded performance data from a group of individuals and, using the maximum likelihood principle, estimate the group-average model parameters and their corresponding variances. Thereafter, as new PVT performance measurements for an individual, not necessarily studied previously, become available, a Bayesian framework makes use of the learned inter-individual variability of the group and the individual's performance data to continuously adapt the model parameters (and the model) to that individual. In contrast, the method proposed by Rajaraman et al.[11] adapts the two-process model parameters for an individual based solely on the individual's performance data, without requiring inter- and intra-individual variability information from a group of individuals. After an initial set of performance data for an individual is collected, as each new performance datum becomes available, the parameters of the two-process model are continuously adapted to that specific individual by solving a constrained, linear least squares (LS) problem, in which the constraint provides a trade-off between the goodness of the model fit and the requirement that the solution follows the two-process model. If a new performance measurement is unavailable at the "expected" sampling period,

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

both methods bypass model adaptation and estimate future performance levels using the most recent parameter values.

Both of these methods have advantages and limitations. The method by Van Dongen et al.[10] enables adaptation of the two-process model parameters using only a few performance measurements from an individual. Moreover, based on the probability distributions of the model-parameter estimates, the adopted Bayesian framework allows for the empirical estimation of 95% confidence intervals around the predictions. Although these empirically computed confidence intervals are not validated through Monte Carlo simulations, and their soundness has been questioned,[17] they provide a first step in addressing a desired capability of performance models that, until now, has been lacking. However, some limitations do exist: Because of the nonlinearity of the two-process performance model with respect to its parameters,[13] the LS solution for finding the best estimates of the parameters involves solving a nonlinear programming (NLP) problem. In such problems, unless the objective function (i.e., the LS equation) is known to be convex, the NLP solution may be a local minimum. Therefore, the proposed Bayesian framework[10,16] cannot guarantee unique (optimal) estimates of the model parameters. This implies that the proposed method may negate a key convergence property of adaptive, or learning, algorithms in that the model-parameters' estimates must converge asymptotically to their underlying true (unknown) values as the number of measurements increases and as the amount of uncertainty in the measurements decreases.[17,18] Furthermore, obtaining the optimal parameter estimates through their proposed Bayesian framework is computationally expensive, involving a brute-force, grid-search procedure in which the computational cost increases as the desired accuracy of the resulting parameter estimate increases.

The method proposed by Rajaraman et al.[11] addresses some of these limitations. It recasts the otherwise nonlinear optimization problem into a set of linear optimization problems, whose solution, if it exists, is guaranteed to be unique. The model-parameter estimates are obtained by solving regularized linear LS problems,[11] avoiding the computationally costly grid-search optimization. More importantly, this method ensures that the model parameters asymptotically converge to their true values as the number of measurements increases and the amount of noise in the observed measurements decreases to zero. Despite these advantages, Rajaraman et al.[11] did not address two shortcomings. First, their method requires the accumulation of a minimum number of past performance measurements of an individual before model-parameter estimates and performance predictions can be made. In theory, at least 13 measurements are required. However, in practice, due to noise in the measured performance data, a larger number of measurements are generally needed. This limitation precludes application of the method to partial/ chronic sleep restriction scenarios, in which, because of intermittent sleep/wake periods, accumulation of large consecutive performance measurements is not possible. Second, they failed to compute any measure of reliability of the model predictions. Such measures would provide statistical error bounds within which model predictions may be trusted for a predefined coverage probability, say, 95%.

In this paper, we extended our previous work by proposing solutions to these two shortcomings. To overcome the requirement that a minimum number of performance data points be collected from an individual before the model can be customized and used to make predictions, we employed a Bayesian-like approach akin to that of Van Dongen.[10] In this approach, we combined *a priori* performance information with the information contained in the individual's measured performance data. As a result, the proposed method leveraged *a priori* performance information and started adapting the model parameters and making predictions for a specific individual as soon as the first performance measurement from that individual became available. However, by retaining the strategy of transforming the nonlinear optimization problem into a series of linear problems, whose solution is the exact solution of the original nonlinear problem, the proposed method guaranteed unique, optimal estimates of the two-process model parameters, avoiding a brute-force, grid-search procedure for computing the estimates. To quantitatively estimate the reliability of the model predictions, we reformulated the two-process model, using its linear representation,[11] into an autoregressive (AR) model framework,[19-22] which directly provides analytical expressions for computing statistically based error bounds around the model predictions in the form of prediction intervals (PIs).

We used two distinct data sets to evaluate the proposed method. The first consisted of simulated performance data on a PVT scale generated from the two-process model with known parameters and superimposed white Gaussian noise. Simulated data allowed us to verify the convergence of the model-parameter estimates to their true values, analyze the trade-off between *a priori* performance information and measured performance data, and qualitatively and quantitatively assess the PIs. We found that the proposed method yielded parameter estimates that asymptotically converged to their true values as the number of performance measurements for an individual increased and the amount of uncertainty (noise) in the measurements decreased. Moreover, the new method yielded improvements in parameter-estimate accuracy of up to 90% over our previous method.[11] In addition, the soundness of the computed PIs was validated through Monte Carlo simulations.

The second data set consisted of PVT lapses obtained from a laboratory study of individuals subjected to 82 h of total sleep loss. The laboratory data allowed testing of the proposed approach within the context of the inter- and intra-individual variability encountered in actual performance data and making comparisons between individualized predictions and group-average model predictions. For individuals representing three distinct sleep-loss phenotypes ("vulnerable," "average," and "resilient"), we found that the new method yielded individualized predictions were up to 43% more accurate than the group-average model predictions and better captured the circadian and homeostatic variation of the performance data. Moreover, prediction accuracy improved, on average, by 10% compared with the previous method.

## METHODS

In this paper, we proposed a new method that employed the two-process model of sleep regulation[12,13] as the underly-

ing model structure to enable individualized prediction of performance impairment due to total sleep loss. This method is based on our previous work,[11] in which, by taking advantage of the linear representation of the two-process model, we estimate unique performance model parameters for an individual after collecting a minimum number of data points. Here, we improved upon the previous work by circumventing the requirement that a minimum number of previously collected data points be available before model-parameter estimation and performance predictions can commence. With this new method, model individualization was started as soon as the first performance measurement became available. This was achieved through Bayesian inference, in which *a priori* performance information was combined with the information available from the individual's measured performance data. The *a priori* performance information was obtained from two-process model predictions generated using *a priori* model-parameter values. As each new performance measurement became available, it was added to the individual's past performance data and together used to update and further individualize the model-parameter estimates. Based on the most recent parameter estimates, we made predictions in accordance with a desired prediction horizon. We took further advantage of the linear representation of the two-process model to reformulate it into an AR model framework, which provides a direct analytical estimation of the reliability of the model predictions in terms of PIs.[19]

## Two-Process Model of Sleep Regulation

The two-process model of sleep regulation consists of two separate processes:[12] Process S (sleep homeostasis), which is dependent on sleep/wake history, increases exponentially with wake time and decreases exponentially with sleep/recovery time to a basal value,[13] whose rates of increase/decrease are individual-specific, assumed to be constant, and have unknown values; and Process C (circadian), which is independent of sleep/wake history and represents a self-sustaining oscillator with a 24-h period.[23] The equations comprising the two-process model at discrete sampling time index $k$ can be expressed as[13,24]

$$S(k) = 1 - \exp(-T_s / \tau_r)(1 - S(k-1)), \text{ during wakefulness,} \quad (1)$$

$$S(k) = \exp(-T_s / \tau_d)(S(k-1)), \text{ during sleep,} \quad (2)$$

$$C(k) = \sum_{i=1}^{5} a_i \sin\left(\frac{i2\pi}{\tau}\left((k-1)T_s + \phi\right)\right), \quad (3)$$

where $S(k)$ denotes the value of the sleep homeostat at time $k$, usually scaled between 0 and 1;[24] $C(k)$ denotes a five-harmonic sinusoidal equation that approximates the circadian oscillator under entrained conditions;[25] $T_s$ represents the sampling period; $\tau_r$ and $\tau_d$ represent the time constants of Process S during wakefulness and sleep, respectively; $\tau$ denotes the time period of the circadian oscillator (~24 h); $a_i$, where $i = 1,...,5$, represents 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 $a_5 = 0.001$); and $\phi$ denotes the initial circadian phase.[23]

As suggested by Achermann and Borbély,[26] we formulated the temporal pattern of cognitive performance as the additive interaction of Processes S and C. Accordingly, performance $P(k)$ at time $k$ was expressed by

$$P(k) = \alpha S(k) + \beta C(k), \quad (4)$$

where $\alpha$ and $\beta$ are real-valued positive parameters that control the trade-off between the two processes of the model. For total sleep deprivation, Eq. (4) can be rewritten as

$$P(k) = \alpha - \alpha S(0)\gamma^{k-1} + \sum_{i=1}^{5} \bar{a}_i \sin\left[i\omega\left((k-1)T_s + \phi\right)\right], \quad (5)$$

where $\gamma = \exp(-\rho T_s)$, with $\rho = 1/\tau_r$, $\omega = 2\pi/\tau$ and $\bar{a}_i = \beta a_i$. Equation (5) is a function of five unknown parameters, $\alpha$, $\gamma$, $\beta$, $S(0)$, and $\varphi$, of which $\alpha$, $\gamma$, and $\beta$ have been termed *trait* parameters, and $S(0)$ and $\varphi$ have been termed *state* parameters.[10,11]

## Individual-Specific Model Development Based on Performance Measurements

In our previous work,[11] we showed that performance $P(k)$ in Eq. (5) is composed of a linear combination of three basic components, a constant signal $\alpha$, an exponential signal $\alpha S(0)\gamma^{k-1}$, and five sinusoidal signals $\bar{a}_i \sin\left[i\omega\left((k-1)T_s + \phi\right)\right]$, where $i = 1,...,5$. Moreover, we note that, if $x(k)$ denotes the value of time series $x$ at time $k$, a shift operator $Z$ can be defined such that $Z^n\{x(k)\} = x(k+n)$. Accordingly, we establish three linear operators, $Z-\gamma$, $Z-1$ and $\prod_{i=1}^{5}(Z^2 - 2\cos(i\omega T_s)Z + 1)$,

so that when each operator is individually applied to performance $P(k)$ in Eq. (5), it eliminates one of the three basic components. Hence, by successively applying the three operators in any order to $P(k)$ in Eq. (5),

$$\left[(Z - \gamma)\left(\prod_{i=1}^{5}(Z^2 - 2\cos(i\omega T_s)Z + 1)\right)(Z - 1)\right]P(k) = 0, \quad (6)$$

we eliminate all three basic components of $P(k)$ and obtain an identically zero equality. The expression within the brackets can be expanded into a 12th-order polynomial in $Z$, which, given the definition of $Z$, represents a 12th-order linear, forward-difference equation. That is, for any set of values for the five parameters $[\alpha, \gamma, \beta, S(0), \text{ and } \varphi]$, $P(k)$ in Eq. (5) is a solution of the 12th-order difference equation represented by Eq. (6).

To adapt these parameters and develop individual-specific models, our previous method only makes use of current and past performance measurements, such as PVT lapses, from the individual being modeled. In this formulation, the relationship between PVT performance measurements $y(k)$, with $k = 1, ..., N$, and $P(k)$ in Eq. (5) with unknown parameter values can be written as

$$y(k) = P(k) + \varepsilon(k), \quad (7)$$

where $\varepsilon(k)$ denotes a normally distributed error (measurement noise) with zero mean and variance $\sigma^2$. We uniquely estimate $P(k)$, or equivalently, the five parameters of the two-process model in Eq. (5) from $y(k)$, by successively applying a pair of the three linear operators to $y(k)$ and solving the three resulting linear constrained LS problems: the first LS problem solves for

$\gamma$, the second for $\alpha$ and $S(0)$, and the third for $\beta$ and $\varphi$.[11] For example, we estimate a unique value for $\gamma$ by solving the following two-step constrained linear LS problem

$$\arg\min_{0 \le \gamma < 1}\left\{\arg\min_{\mathbf{P}}\left[\|\mathbf{P} - \mathbf{y}\|_2^2 + \lambda^2\|\mathbf{L}_\gamma\mathbf{P}\|_2^2\right]\right\}, \qquad (8)$$

where $\mathbf{y}$ denotes the $N$ x 1 vector of performance measurements $y(k)$, with $k = 1, \ldots, N$; $\mathbf{P}$ represents the corresponding $N$ x 1 vector of the performance fit $P(k)$; $\mathbf{L}_\gamma$ denotes the $N$-12 x $N$ matrix that represents the expression within the brackets in Eq. (6); and $\lambda^2$ is a positive real number. To minimize Eq. (8), we first fix the value of $\gamma$ ($0 \le \gamma < 1$) and, through constrained LS,[11] find $\mathbf{P}$ that minimizes the expression within the brackets, for a chosen $\lambda^2$. In this formulation, $\lambda^2$ provides a trade-off for the solution of $\mathbf{P}$ between fitting the performance measurements $\mathbf{y}$ (the first term inside the brackets) and satisfying the constraint imposed by $L_\gamma$ (the second term),[27] which requires that $\mathbf{P}$ follows the two-process model in Eq. (5). By setting $\lambda^2$ to an arbitrarily large value, we forced the solution of $\mathbf{P}$ to follow the two-process model, while simultaneously fitting the measurements $\mathbf{y}$. To avoid potential numerical problems with this procedure, we transformed Eq. (8) to its standard form,[28] solved it with $\lambda^2$ set to an arbitrarily large value, and transformed its solution back to the original problem. Next, we change the value of $\gamma$ and repeat this process over the entire range of possible values, $0 \le \gamma < 1$, to estimate the *optimal* $\gamma$, which minimizes Eq. (8). Using the optimal $\gamma$ and the corresponding solution of $\mathbf{P}$ from Eq. (8), we estimate the other four parameters of the two-process model by solving two additional constrained linear LS problems.[11] As new performance measurements from the individual being modeled become available, they are combined with the previous measurements $\mathbf{y}$ and together used to adapt the model parameters for that individual by repeating the above procedure.

A limitation of this method is the requirement that a minimum number of past performance measurements be available for the individual before model-parameter estimation and performance prediction can begin. In theory, at least 13 measurements are required; however, in practice, due to noise in the measurements $\mathbf{y}$, a larger number of measurements is generally needed.

### Individual-Specific Model Development Based on A priori Knowledge and Performance Measurements

To overcome this limitation and permit model individualization and performance prediction as soon as the first measured performance datum becomes available, we modified the above approach by considering *a priori* performance information in addition to the information available from the individual's measured performance data. The *a priori* performance information was obtained from performance data generated from the two-process model in Eq. (5), with its model parameters set to *a priori* estimated values. This Bayesian approach was mathematically implemented by adding a term to Eq. (8), which accounted for $M$ *a priori* data $\tilde{y}(k) = 1\text{-}M, 2\text{-}M,\ldots,0$, and solving the augmented expression through the series of optimizations

discussed above. For example, $\gamma$ was estimated by solving the two-step constrained linear LS problem

$$\arg\min_{0 \le \gamma < 1}\left\{\arg\min_{\overline{\mathbf{P}}}\left[\|\mathbf{P} - \mathbf{y}\|_2^2 + \mu^2\|\widetilde{\mathbf{P}} - \widetilde{\mathbf{y}}\|_2^2 + \lambda^2\|\mathbf{L}_\gamma\overline{\mathbf{P}}\|_2^2\right]\right\}, \qquad (9)$$

where $\tilde{y}$ denotes the $M$ x 1 vector of prior performance data $\tilde{y}(k)$; $\mu$ is a positive real number; and $\overline{\mathbf{P}}$ represents the $(N + M)$ x 1 vector of the performance fit $\overline{P}(k)$, with $k = 1\text{-}M, 2\text{-}M,\ldots, N$, whose first $M$ elements are represented by $\widetilde{\mathbf{P}}$ and the remaining $N$ elements are represented by $\mathbf{P}$. In other words, $\overline{\mathbf{P}}$ denotes the concatenation of vectors $\widetilde{\mathbf{P}}$ and $\mathbf{P}$.

The first term in Eq. (9) quantifies the fit of the solution of $\overline{\mathbf{P}}$ to the individual's measured performance data $\mathbf{y}$, whereas the second term quantifies the fit to *a priori* generated performance data $\tilde{y}$, with degree $\mu^2$. Accordingly, $\mu^2$ provides a mechanism to obtain a solution of $\overline{\mathbf{P}}$ that provides a trade-off between our trust in the measured performance data and the *a priori* performance information. As $\mu^2$ decreases, we increase our trust in the measured data and emphasize the fit of $\overline{\mathbf{P}}$ to $\mathbf{y}$, giving less weight to the prior information. This shift in emphasis increases the variance of the solution of $\overline{\mathbf{P}}$ (i.e., it increases the trace of the covariance of $\overline{\mathbf{P}}$), since the $\mathbf{y}$ measurements are noisy.[29] In the limit where $\mu^2 = 0.0$, the *a priori* performance information is not considered, and the last $N$ elements of the solution of $\overline{\mathbf{P}}$ in Eq. (9) are identical to $\mathbf{P}$ obtained from solving Eq. (8). Conversely, as $\mu^2$ increases, we increase our trust in the prior information and the model-parameter estimates obtained from the solution of $\overline{\mathbf{P}}$ converge to the *a priori* selected values.

The optimal value for $\mu^2$ was computed by the algorithm described in Appendix A and is a function of a user-provided estimate of the uncertainty (i.e., noise level) $\hat{\sigma}^2$ in the measured performance data $\mathbf{y}$. The computed $\mu^2$ was optimal in the sense that, given $\hat{\sigma}^2$, it minimized a cost function that simultaneously accounted for the fit to the measured data and the fit to the prior information. Accordingly, as the provided uncertainty in the measured data $\hat{\sigma}^2$ decreased, the optimal $\mu^2$ decreased, accentuating the trust in the measured data, and vice versa. The algorithm for the optimal selection of $\mu^2$ endowed another desired feature of Bayesian estimation algorithms[30] in that the trust in the measurements (i.e., observations) outweighs the trust in the *a priori* information as the number of measurements provided to the algorithm increases. As a result, the optimal $\mu^2$ decreased asymptotically as the number of performance measurements $N$ from an individual increased. That is, as more and more performance measurements were attained for an individual, the algorithm placed more and more trust in the measurements, de-emphasizing the prior information. The rate of decrease of $\mu^2$, however, became faster or slower as the user-provided uncertainty $\hat{\sigma}^2$ in the measurements $\mathbf{y}$ decreased or increased, respectively.

We solved Eq. (9) for the *optimal* estimate of $\gamma$ following steps similar to those used in solving Eq. (8) discussed above. We performed this procedure over the entire range $0 \le \gamma < 1$ to estimate the *optimal* $\gamma$. Next, using the optimal $\gamma$ and the corresponding solution of $\overline{\mathbf{P}}$, we uniquely estimated the other four parameters of the two-process model by solving the associated constrained LS problems. As new performance measurements from the individual became available, they were combined with the previous measurements $\mathbf{y}$ and used together with the prior

information ỹ to adapt the model-parameter estimates and make new predictions.

<u>**Prediction Uncertainty Quantification**</u>

Under certain conditions, individualized predictions may be of limited value unless they are accompanied by measures of reliability of prediction in the form of statistically based error bounds, such as PIs.[31] One way to estimate these PIs within the context of the two-process model is to first compute *confidence* intervals for the model-parameter estimates and then translate them into PIs through Eq. (5). However, the nonlinear relationship between the two-process model and its parameters[13] precludes the derivation of analytical functions that map confidence intervals of the parameters' estimates onto PIs for the model predictions.[32] Rather, PIs are often estimated through computationally expensive Monte Carlo simulations.[33]

Here, we analytically computed PIs about the model predictions by taking advantage of the linear representation of the two-process model in Eq. (5). Using the property that $P(k)$ in Eq. (5) is a solution of the 12th-order difference equation in Eq. (6),[11] we considered $P(k)$ as a 12th-order autoregressive (AR) process (see Appendix B) in which, for each time $k$, $P(k)$ is expressed as a linear combination of previous performance measurements,[19, 22, 34] such that

$$P(k) = \sum_{i=1}^{12} b_i P(k-i), \qquad (10)$$

where $b_i$, with $i = 1,\dots,12$, denotes the AR-model coefficients. By considering the correspondence between Eqs. (10) and (6) [see Appendix B], we obtained $b_i$ as the coefficients of the 12th-order polynomial given by the expression within the brackets in Eq. (6). Finally, by taking advantage of the availability of an analytic expression to compute statistically based error bounds for AR-model predictions,[19] we obtained the following equation for computing PIs with a coverage probability of $100(1-\theta)$ %

$$PI_{100(1-\theta)\%}(k) = P(k) \pm Z_{\theta/2} \sqrt{\mathbf{b} \Sigma \mathbf{b}^T + \hat{\sigma}^2}, \qquad (11)$$

where $\theta$ represents the significance level; $Z_{\theta/2}$ represents the percentage point of a standard normal distribution with a proportion $\theta/2$ above it; $\mathbf{b}$ denotes the 1 x 12 vector of coefficients $b_i$, with $i = 1,\dots,12$; $P(k)$ denotes the performance prediction at time $k$ given previous measurements $y(k-1), y(k-2),\dots$; $\Sigma$ denotes the covariance matrix of $\overline{P}(k-1),\dots,\overline{P}(k-12)$, obtained by solving Eq. (9); and $\hat{\sigma}^2$ denotes the user-provided noise-level estimate of the performance measurements $\mathbf{y}$. We computed $P(k)$ through Eq. (10), where $P(k-i)$, $i = 1,\dots,12$, were obtained from the solution of $\overline{\mathbf{P}}$ in Eq. (9). To compute $PI_{100(1-\theta)\%}(k-n)$, for $n \geq 1$, we recursively applied Eq. (10) to generate predictions at $k + m$, with $m = 1,2,\dots,n$, followed by Eq. (11) for computing the corresponding PIs.

## RESULTS

We employed two data sets to validate our proposed approach. The first consisted of simulated performance data obtained from running the two-process performance model in Eq.

(5) with its parameters set to selected values and superimposing random noise to emulate the uncertainty present in measured performance data. Simulated data enabled verification of key convergence properties of the proposed approach, sensitivity analyses of the trade-off between prior information and measured data, and quantitative and qualitative assessment of the analytically computed PIs.

The second data set consisted of PVT lapse measurements from nine healthy individuals who had participated in a 82-h total sleep deprivation study under laboratory conditions.[11,35] The laboratory data set allowed testing of the proposed approach within the context of the inter- and intra-individual variability that characterizes measured performance data and comparing the individualized predictions versus those obtained with group-average prediction models.

**Simulated Data Set**

We generated simulated data sets by running the two-process performance model in Eq. (5) with known parameters and superimposing white (i.e., uncorrelated) Gaussian noise to emulate the variability observed in measured performance data. By sampling from the group-average distribution of parameters estimated by Van Dongen et al.,[10] we set the three trait parameters in Eq. (5) to $\alpha = 30.30$, $\beta = 6.35$, and $\rho = 0.03$ h$^{-1}$ ($\gamma = 0.94$), whereas by sampling from uniform distributions with support [0 1] and [0 24], respectively, we set the state parameters S(0) = 0.82 and $\varphi = 6$ h. In each simulation, performance data were generated for 82 h of total sleep deprivation and sampled once every 2 h to emulate the sampling rate of performance measurements in typical sleep studies. Data from each simulation were superimposed with white noise sampled from a Gaussian distribution with zero mean and one of three levels of variance: 16, 4, or 1. These variances at 82 h of wakefulness corresponded to a signal-to-noise ratio (SNR) of 3.70 (59.10/16), 14.77 (59.10/4), and 59.10 (59.10/1), respectively, where 59.10 was the variance of the performance data generated by running Eq. (5) with the abovementioned parameters for 82 h. The SNR is defined as the ratio of the variance of a signal to the variance of the white noise corrupting that signal. In each simulation, the *a priori* values of the trait parameters, $\alpha_{pr} = 29.70$, $\beta_{pr} = 4.30$, and $\rho_{pr} = 0.03$ h$^{-1}$ ($\gamma_{pr} = 0.94$), were chosen to match Van Dongen's estimates of their group-average model parameters,[10] whereas the *a priori* values of the state parameters, S(0)$_{pr}$ = 0.92 and $\varphi_{pr} = 12.6$ h, were randomly chosen by sampling from uniform distributions, as above. These same parameters were used in the group-average model predictions for all simulations. Based on these values, we employed Eq. (5) to generate 13 prior performance data points. Thirteen is the minimum number of data points required to recover the two-process model-parameter values, since the model is the solution of the 12th-order difference equation [i.e., Eq. (6)]. Although a greater amount of previously collected data may be used, we empirically found that 13 data points provided the fastest rate of convergence of the model-parameter estimates to their true values.

To make the first prediction, the 13 *a priori* data were combined with the first performance measurement $y(1)$ to estimate the homeostatic-rate parameter $\rho$, or equivalently $\gamma$, by minimizing Eq. (9). We set $\lambda^2$ to $2^{1024}$ (the largest-possible double
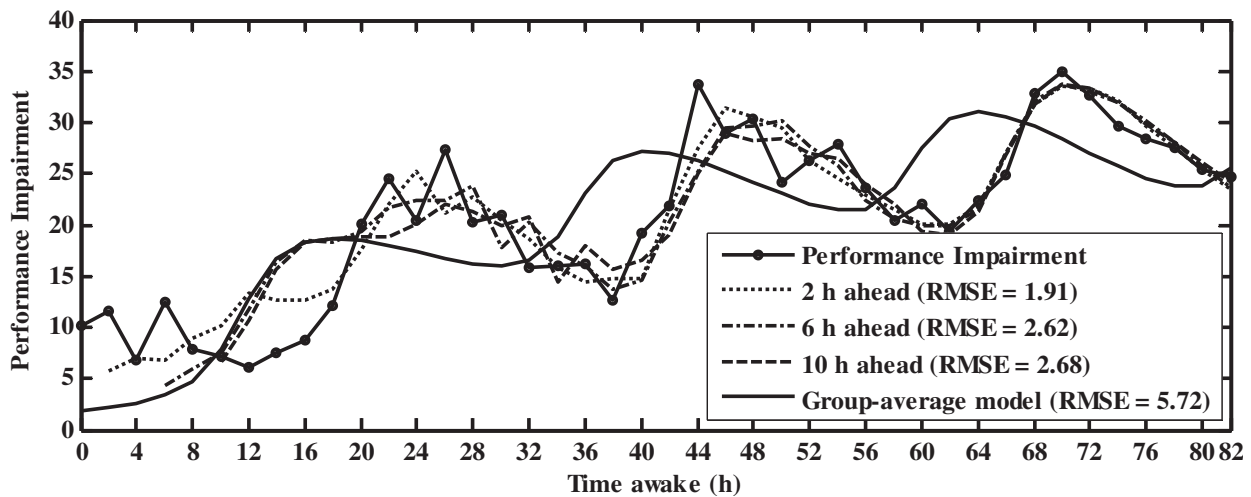
**Figure 1**—Simulated performance data, normalized to a psychomotor vigilance task scale, for 82 h of total sleep deprivation with a signal-to-noise ratio of 14.77. The solid circles show the simulated performance lapses every 2 h. The dotted, dash-dotted, and dashed lines represent 2-, 6-, and 10-h-ahead predictions, respectively, based on individualized models with prior parameter values as those used in the group-average model and noise-level estimate set to the correct value ($\sigma^2 = 4$). The solid line represents the group-average model predictions. We used the root mean squared error between the predicted and the simulated performance data, calculated from 10 through 82 h, as a comparative metric.

precision floating point number) and, given the user-provided uncertainty $\hat{\sigma}^2$ in $y(1)$, estimated the optimal $\mu^2$ as discussed in the Methods Section. The remaining four parameters of the two-process model [$\alpha$, $S(0)$, $\beta$, and $\varphi$], were estimated by solving the associated LS problems,[11] with $\overline{\mathbf{P}}$ taken as the solution of Eq. (9). Thereafter, as new performance measurements became available, they were added to previous measurements, one at a time, and, by following the procedure above, the model parameters were adapted to take into account the entire set of measurements up to the last datum. Using the most recent estimates of the parameters, we made performance predictions for a desired prediction horizon. For example, a 6-h-ahead prediction at time index $k$ was made using the model-parameter estimates obtained at time index $k$-3.

Figure 1 shows a representative realization of the performance profile of a simulated individual, with SNR = 14.77 at 82 h of wakefulness (i.e., $\sigma^2 = 4$), and individualized performance predictions for three selected prediction horizons of 2, 6, and 10 h. The performance units were normalized to emulate results of typical 10-min PVT laboratory studies. The root mean squared error (RMSE) was used as a metric to quantify the accuracy of the predictions (the smaller the RMSE between the predicted and the true underlying performance signal, the better is the resulting prediction).[19] For consistency, RMSEs were computed across overlapping prediction time intervals spanning from 10 to 82 h. As shown in Figure 1, the individualized model was consistently more accurate than the group-average model, yielding a two- to three-fold reduction in prediction errors. Moreover, we note that the group-average model predictions (solid line) were out of phase with the simulated performance profile (solid circles). As expected, the RMSEs increased monotonically with increasing prediction horizons. This should be expected because the model parameters were not updated over the prediction horizon, preventing longer-horizon predictions from employing intermediate measurements to adapt the model parameters and improve the predictions. However, it should be noted that uti-

lization of intermediate measurements with low SNR could result in incorrect model-parameter estimates that, while yielding a good fit to the measurements, result in poorer predictions.[18] As a result, low SNR measurements could yield predictions over a longer horizon with smaller RMSEs than ones over shorter horizons. Nonetheless, as the number of measurements increases, the model parameters estimates should converge to their true (unknown) values, and the RMSEs should be approximately the same regardless of the prediction horizon.[11]

For the simulation in Figure 1, Figure 2 shows the parameter estimates (solid squares), their true values (dashed lines), and *a priori* values (solid lines) of the five model parameters as a function of the number of available performance measurements, starting with the first PVT measurement at time zero. At first, the parameter estimates oscillate due to the low SNR in the simulated performance data during this initial period. Later, as additional measurements became available and the variance in the data increased, achieving a SNR = 14.77 at 82 h of wakefulness, each of the parameters asymptotically converged to their true values, attaining 95% accuracy after ~50 h of wakefulness.

To verify the ability of the algorithm to consistently improve the estimation of the model parameters as a function of data uncertainty $\hat{\sigma}^2$ and the number of performance measurements $N$, we performed three sets of simulations. Each consisted of 100 trials, in which each trial in a set was superimposed with white noise sampled from one of the three Gaussian distributions with mean zero and variances that, at 82 h of wakefulness, corresponded to SNRs of 59.10 ($\sigma^2 = 1$), 14.77 ($\sigma^2 = 4$), and 3.70 ($\sigma^2 = 16$). In this way, we observed the impact of each noise level on parameter estimation. Table 1 shows the mean, the variance, and the mean squared error (MSE), i.e., the square of the bias plus variance, for each of the five parameter estimates over the 100 trials for the three sets of simulations. The MSE, which accounts for both estimation accuracy (bias) and estimation precision (variance), is the most statistically meaningful
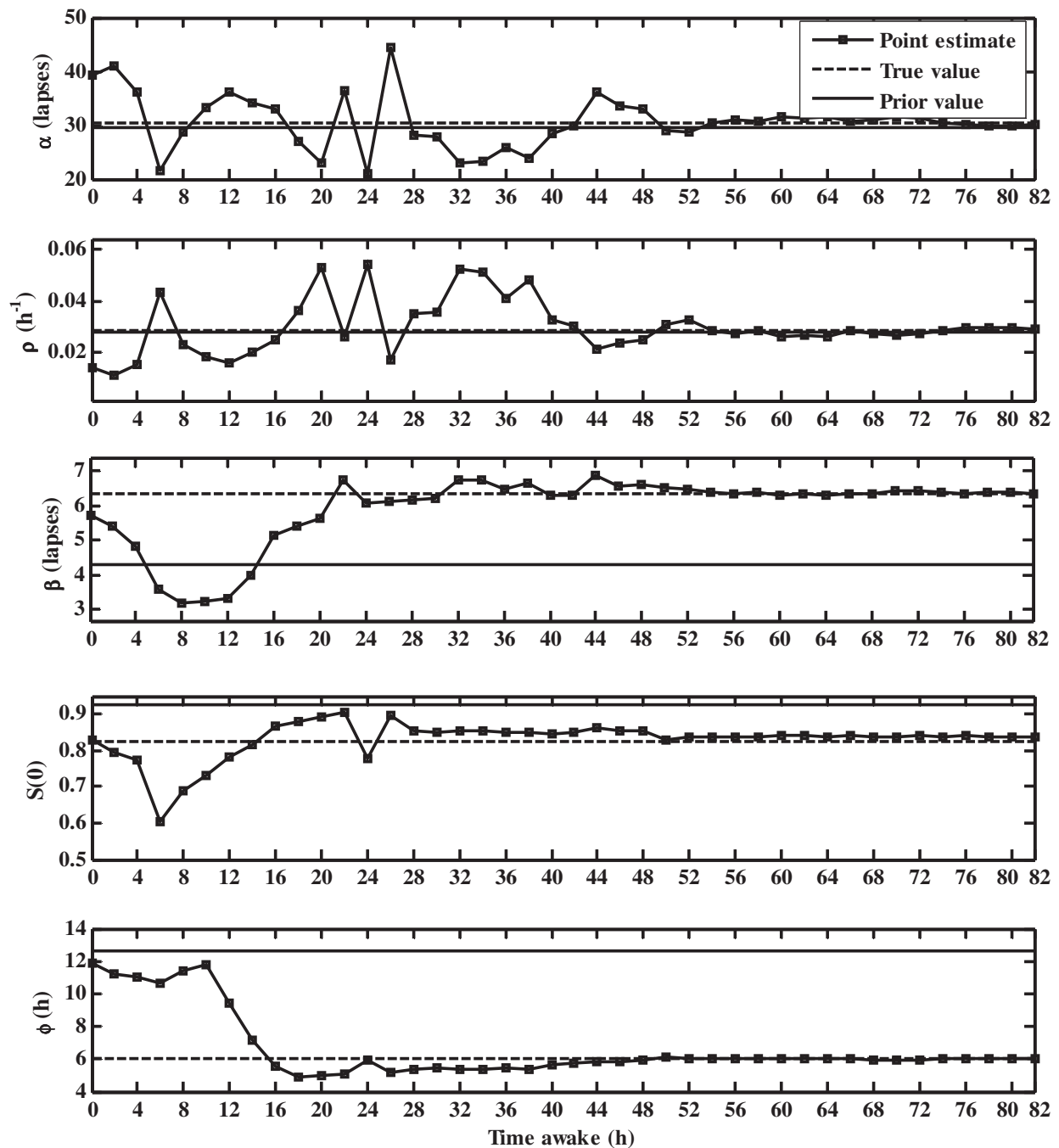
**Figure 2**—Estimates (solid squares), the true value (dashed lines), and the prior values (solid lines) of the five parameters of the two-process model. These correspond to the simulated performance data in Figure 1 and show that the model-parameter estimates asymptotically approached their true values as they were adjusted every 2 hours, as new measurements became available.

metric,[36] and can be quantified in simulated data where the bias in the estimate can be determined. The parameter estimates are shown at four points in time, at 22, 42, 62, and 82 h of wakefulness, for each of the three SNRs. For each of the parameters, the results consistently indicated that, for a fixed time point, the MSE increased as the SNR decreased and, for a fixed SNR, the MSE decreased as the number of available performance measurements increased. The only exception was for $\alpha$ at 22 h of wakefulness, when the MSE for SNR = 59.10 (104.44) was larger than the corresponding value for SNR = 14.77 (98.49).

This anomaly usually occurs during the early phases of model-parameter estimations and when the SNR is large. In this case, the algorithm de-emphasizes the prior performance information and only trusts the measured data, which do not contain sufficient information for model-parameter estimation. As a result, the estimates have large uncertainty, occasionally causing the MSE to be larger than expected.

We also compared the results from our previous work[11] with the present results. For the particular representative realization of the simulated performance profile in Figure 1, we

**Table 1**—Parameter Estimates of One Individual for Three Monte Carlo Simulations of 100 Trials, Each with Three Different Signal-to-Noise Ratios (59.10, 14.77, and 3.70)

| | | $\alpha$ (lapses) 30.30 | $\rho$ (h$^{-1}$) 0.03 | $\beta$ (lapses) 6.35 | $S(0)$ 0.82 | $\varphi$ (h) 6.00 |
|---|---|---|---|---|---|---|
| **True values** | | | | | | |
| **Signal-to-noise-ratio** | **Time awake (h)** | | | | | |
| 59.10 | | | | | | |
| | 22 | 29.02 (SD = 10.14) MSE = 104.44 | 0.04 (SD = 0.03) MSE = 0.00 | 6.39 (SD = 0.50) MSE = 0.25 | 0.81 (SD = 0.06) MSE = 0.00 | 6.12 (SD = 0.33) MSE = 0.12 |
| | 42 | 30.43 (SD = 3.88) MSE = 15.10 | 0.03 (SD = 0.01) MSE = 0.00 | 6.25 (SD = 0.33) MSE = 0.12 | 0.83 (SD = 0.02) MSE = 0.00 | 6.02 (SD = 0.22) MSE = 0.05 |
| | 62 | 30.29 (SD = 1.50) MSE = 2.25 | 0.03 (SD = 0.00) MSE = 0.00 | 6.29 (SD = 0.27) MSE = 0.08 | 0.83 (SD = 0.02) MSE = 0.00 | 6.01 (SD = 0.18) MSE = 0.03 |
| | 82 | 30.20 (SD = 0.77) MSE = 0.60 | 0.03 (SD = 0.00) MSE = 0.00 | 6.29 (SD = 0.24) MSE = 0.06 | 0.83 (SD = 0.02) MSE = 0.00 | 5.99 (SD = 0.14) MSE = 0.02 |
| 14.77 | | | | | | |
| | 22 | 28.48 (SD = 9.76) MSE = 98.49 | 0.05 (SD = 0.08) MSE = 0.01 | 6.36 (SD = 0.90) MSE = 0.81 | 0.82 (SD = 0.09) MSE = 0.01 | 6.19 (SD = 0.60) MSE = 0.40 |
| | 42 | 30.28 (SD = 5.91) MSE = 34.98 | 0.03 (SD = 0.01) MSE = 0.00 | 6.06 (SD = 0.67) MSE = 0.53 | 0.83 (SD = 0.04) MSE = 0.00 | 6.07 (SD = 0.44) MSE = 0.20 |
| | 62 | 30.26 (SD = 2.63) MSE = 6.93 | 0.03 (SD = 0.01) MSE = 0.00 | 6.20 (SD = 0.54) MSE = 0.32 | 0.83 (SD = 0.03) MSE = 0.00 | 6.03 (SD = 0.35) MSE = 0.13 |
| | 82 | 30.14 (SD = 1.44) MSE = 2.11 | 0.03 (SD = 0.00) MSE = 0.00 | 6.19 (SD = 0.47) MSE = 0.25 | 0.83 (SD = 0.03) MSE = 0.00 | 5.99 (SD = 0.28) MSE = 0.08 |
| 3.70 | | | | | | |
| | 22 | 26.63 (SD = 10.19) MSE = 117.33 | 0.06 (SD = 0.05) MSE = 0.00 | 6.21 (SD = 1.69) MSE = 2.87 | 0.83 (SD = 0.15) MSE = 0.02 | 6.41 (SD = 1.15) MSE = 1.49 |
| | 42 | 28.66 (SD = 5.88) MSE = 37.24 | 0.04 (SD = 0.02) MSE = 0.00 | 5.56 (SD = 1.30) MSE = 2.31 | 0.85 (SD = 0.09) MSE = 0.01 | 6.28 (SD = 0.91) MSE = 0.90 |
| | 62 | 29.81 (SD = 3.44) MSE = 12.09 | 0.03 (SD = 0.01) MSE = 0.00 | 5.88 (SD = 1.09) MSE = 1.41 | 0.85 (SD = 0.07) MSE = 0.01 | 6.18 (SD = 0.77) MSE = 0.62 |
| | 82 | 29.93 (SD = 2.21) MSE = 5.03 | 0.03 (SD = 0.01) MSE = 0.00 | 5.89 (SD = 0.95) MSE = 1.12 | 0.84 (SD = 0.06) MSE = 0.00 | 6.05 (SD = 0.59) MSE = 0.35 |

The estimates' mean, standard deviation (SD), and mean squared error (MSE), i.e., the square of bias plus variance, are shown at four time points: 22, 42, 62, and 82 h of wakefulness.

observed significant prediction and parameter-estimate improvements with the current method. In particular, the average prediction RMSE over the three horizons, computed between 52 and 82 h of wakefulness, was 80% smaller with the current method. The average MSEs over the five parameters shown in Table 1 at 62 h of wakefulness were 50%, 90%, and 93% smaller with the new method when the SNR was set at 59.10, 14.77, and 3.70, respectively, whereas at 82 h of wakefulness, the corresponding values were 45%, 50%, and 89% smaller with the new method. These two time points (62 and 82 h of wakefulness) were the only common points to both studies. The simulation results indicated that the proposed method's improvements in predictions and parameter estimations were more significant when $\hat{\sigma}^2$ in the performance data was high and the number of available measurements for parameter estimation was small.

Figure 3 shows the performance profile, the 10-h-ahead predictions, and their corresponding 95% PIs for the simulated individual in Figure 1with SNRs set at 3.70 (top), 14.77 (middle), and 59.10 (bottom). The 10-h-ahead predictions at any sample index $k$ were actually computed at sample index $k$-5 and are illustrated in all three panels by dashed lines. The dotted lines represent the corresponding upper and lower limits of the

analytically computed 95% PIs. To corroborate these estimations, we empirically computed the 95% PIs (dash-dotted lines) through 100 Monte Carlo simulations. The results indicated that while the analytically computed 95% PIs were initially underestimated, for each of the three SNRs, they converged to the corresponding empirically calculated values as the number of performance measurements increased. This is attributed to the Bayesian aspect of our method, where during the early phases of the model-parameter estimates the algorithm places greater trust on the prior performance information than the measured data, resulting in PIs that are narrower than expected. However, as the number of performance measurements for an individual increased, the 95% PI converged to the corresponding Monte Carlo simulated values, covering almost entirely the individual's performance measurements. The results also indicated that, as expected, the width of the analytically computed 95% PIs decreased from the top to the bottom panels as the SNR in the simulated performance data increased.

To analyze the sensitivity of the *a priori* chosen values of the two-process model parameters on the performance prediction, we compared and contrasted predictions for three simulated individuals, each representing a different sleep-loss phenotype. Figure 4 shows simulated performance profiles (solid circles),
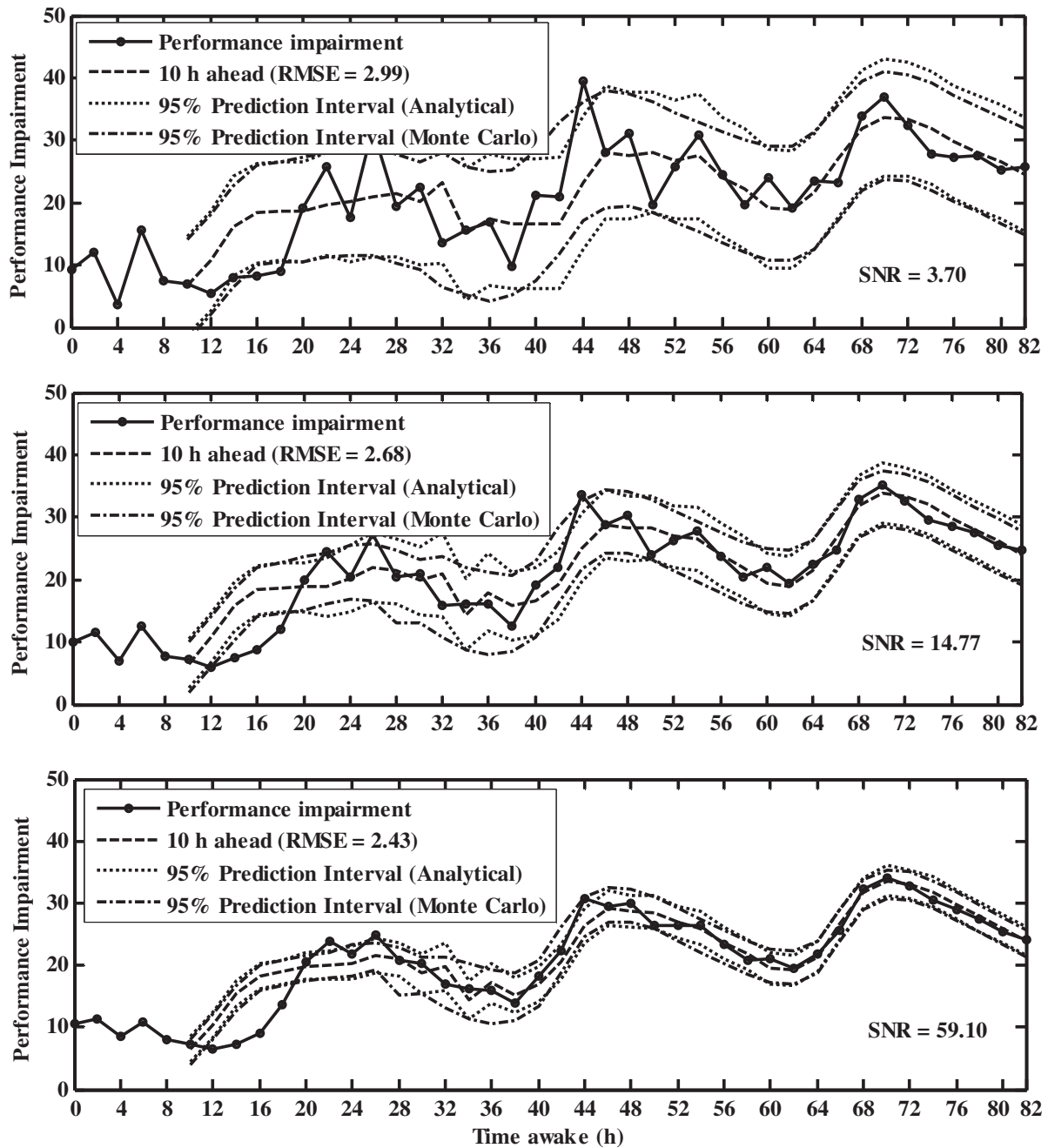
**Figure 3**—Ten-hour-ahead performance predictions (dashed lines) and their corresponding analytically computed 95% prediction intervals (dotted lines) for the simulated performance data shown in Figure 1, with signal-to-noise ratio (SNR) set at 3.70 (top), 14.77 (middle), and 59.10 (bottom). The dash-dotted lines in each of the three panels represent the 95% prediction intervals computed through 100 Monte Carlo simulations.

10-h-ahead predictions (dashed lines), and group-average model predictions (solid lines) for representative "average" (top), "vulnerable" (middle), and "resilient" (bottom) individuals with SNRs set to 14.77. We computed the group-average predictions using the same parameter values as those used for generating *a priori* performance data. The figure shows that the accuracy in the 10-h-ahead predictions, measured in terms of RMSE, varied between individuals, in that the larger the RMSE between the group-average model predictions and the true underlying performance of an individual, the larger the error in the individualized performance predictions for that individual. In other words, as expected, the closer the prior parameter values were

to an individual's true (unknown) parameters, the more accurate the individualized predictions were for that individual.

Except in simulations, one does not know the noise level $\sigma^2$ in performance data. Hence, to analyze the robustness of the proposed individualized prediction method against errors in noise-level estimates $\hat{\sigma}^2$, we compared the predictions for a simulated individual using three different $\hat{\sigma}^2$. Figure 5 shows three 10-h-ahead predictions for the "average" individual in Figure 4 with SNR equal to 14.77. In each prediction, the noise-level estimate $\hat{\sigma}^2$ was set to one-hundredth (dotted), equal to (dashed), and 100 times (dash-dotted) the true noise level ($\sigma^2 = 3.38$) used in simulating performance profile.
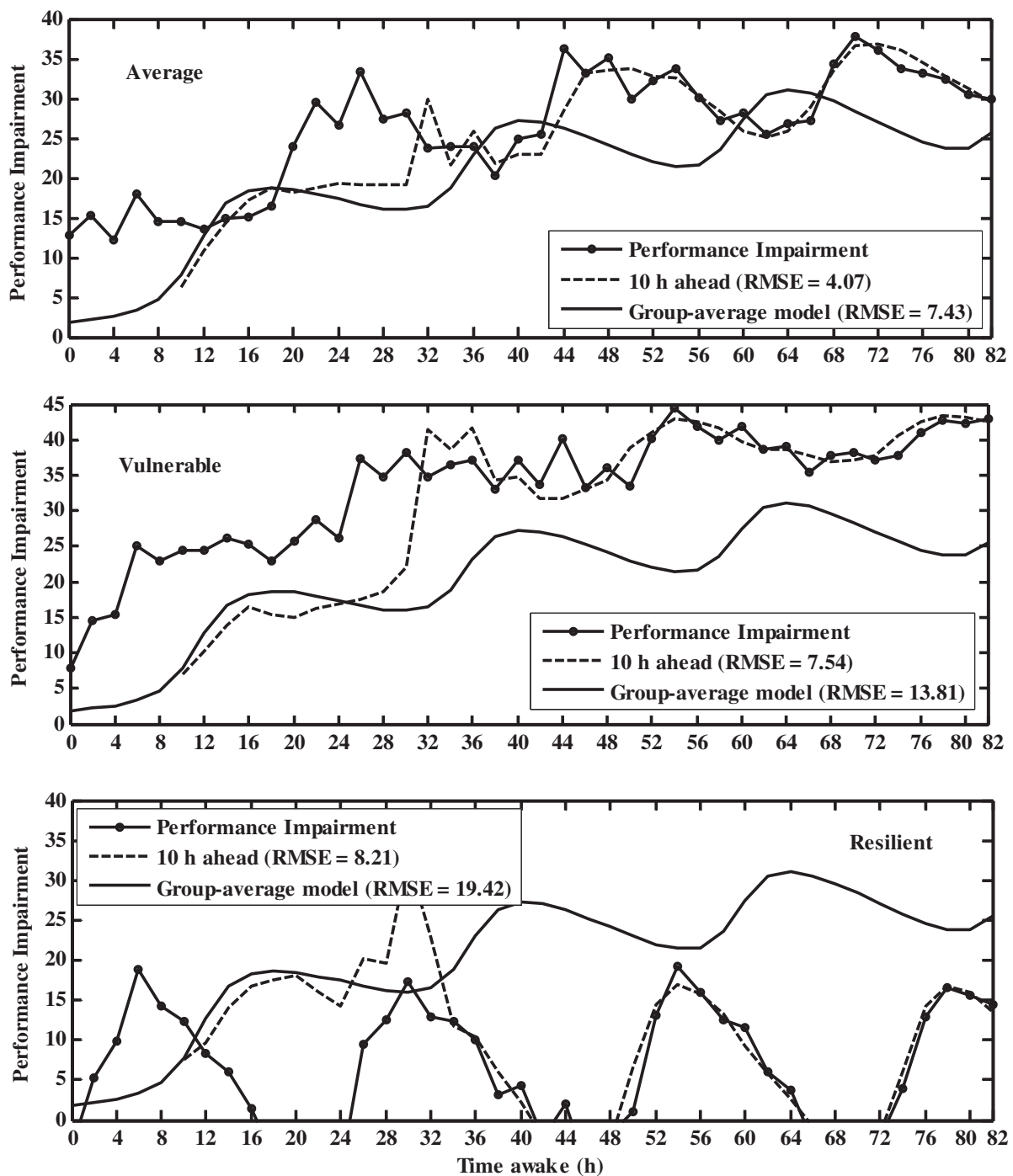
**Figure 4**—Individualized model predictions for three simulated individuals, each representing three different sleep-loss phenotypes [average sensitivity to sleep loss (top), vulnerable to sleep loss (middle), and resilient to sleep loss (bottom)]. The solid circles in each of the panels represent the simulated performance data with a signal-to-noise ratio of 14.77. The dashed lines represent 10-h-ahead predictions using the group-average model parameter values as the prior and using the correct noise level. The solid lines in each panel represent the group-average model predictions.

Inherently, the proposed algorithm followed a key property of learning algorithms,[18] assigning greater trust to the measured performance data as the user-provided noise-level estimate ($\hat{\sigma}^2$) decreased to zero, and vice versa. Figure 5 shows that, while over-estimation of the noise level results in predictions closer to the group-average model predictions, underestimation results in predictions with RMSEs equivalent to those obtained using the correct noise-level estimate. This was because the two-process model is composed of sinusoidal and exponential components, which do not possess enough flexibility to yield performance signals that exactly fit a set of noisy performance measurements. Hence, underestimation of the noise level, which increased the trust on the measured performance data, did not allow the algorithm to overfit the performance data. Although, for a particular
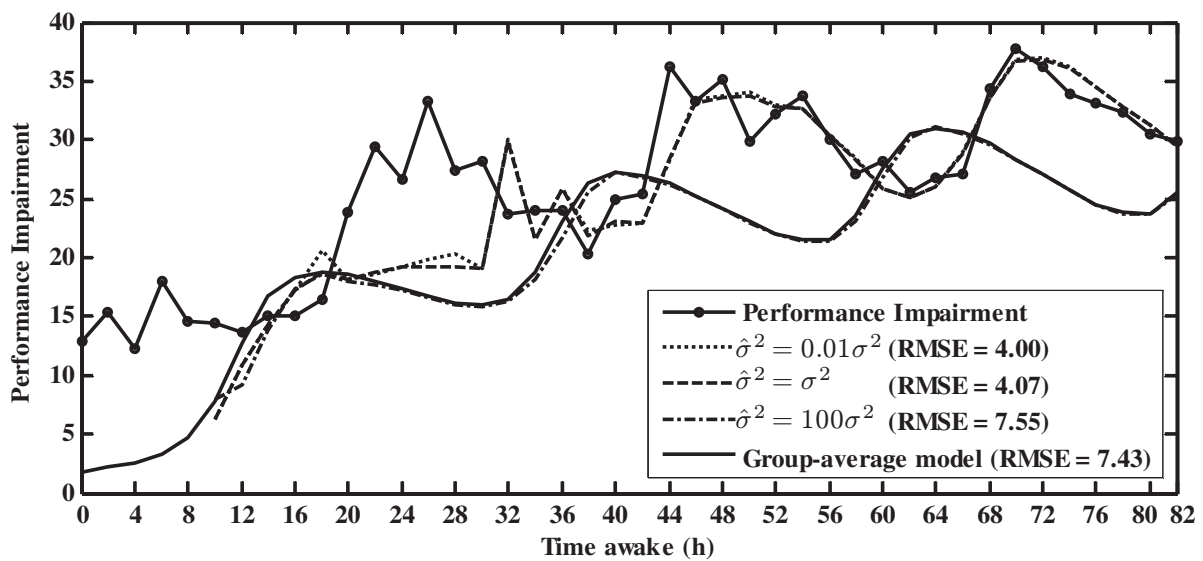
**Figure 5**—Individualized model predictions for the average sleep-sensitivity phenotype in Figure 4 (top). The solid circles represent the simulated performance data with a signal-to-noise ratio of 14.77. The dotted, dashed, and dash-dotted lines represent the 10-h-ahead predictions based on noise-level estimates $\hat{\sigma}^2$ set to one-hundredth, equal to, and 100 times, respectively, the true noise level ($\sigma^2 = 3.38$). The prior parameter values were the same as those used in the group-average model.
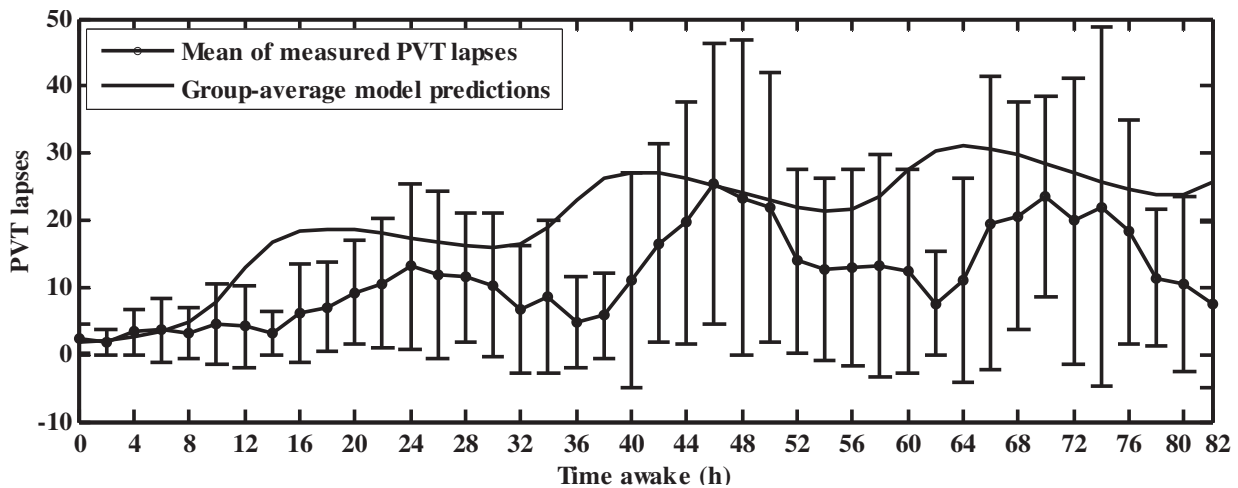


**Figure 6**—Mean and standard deviation of psychomotor vigilance task (PVT) lapse measurements collected every 2 hours for 11 individuals during 82 h of total sleep deprivation. The solid line shows the performance predictions obtained with the group-average model.

realization of noise in the performance data, predictions with under-estimation of the noise level may yield smaller RMSEs than those with the correct noise-level estimate, predictions averaged over a number of noise realizations should yield smaller RMSEs when $\hat{\sigma}^2$ is closer to the true noise level in the measured data.

### Laboratory Data Set

The second data set used to test the proposed approach was collected from a controlled laboratory study in which 48 healthy adults were kept awake for 85 consecutive hours to determine the effects of fatigue countermeasures on performance and alertness during prolonged sleep deprivation.[35] In the present study, we tested a subset of 11 subjects who had not been administered active fatigue countermeasures, and for whom PVT measures were collected every 2 h, for a total of 42 measurements. Figure 6 shows the mean performance profile (solid circles) and the standard deviation of the 11 subjects over the 82 h of PVT data collection and the group-average model predictions (solid lines). Overall, the trend suggested that both PVT lapses and variance (70% of which could be attributable to inter-individual variability[5]) increased over time. Moreover, the figure shows that the group-average model does not even predict the mean-group performance well.

For comparison purposes, we selected the same three subjects as in the previous study,[11] each representing one of three different sleep-loss phenotypes: relatively vulnerable to sleep
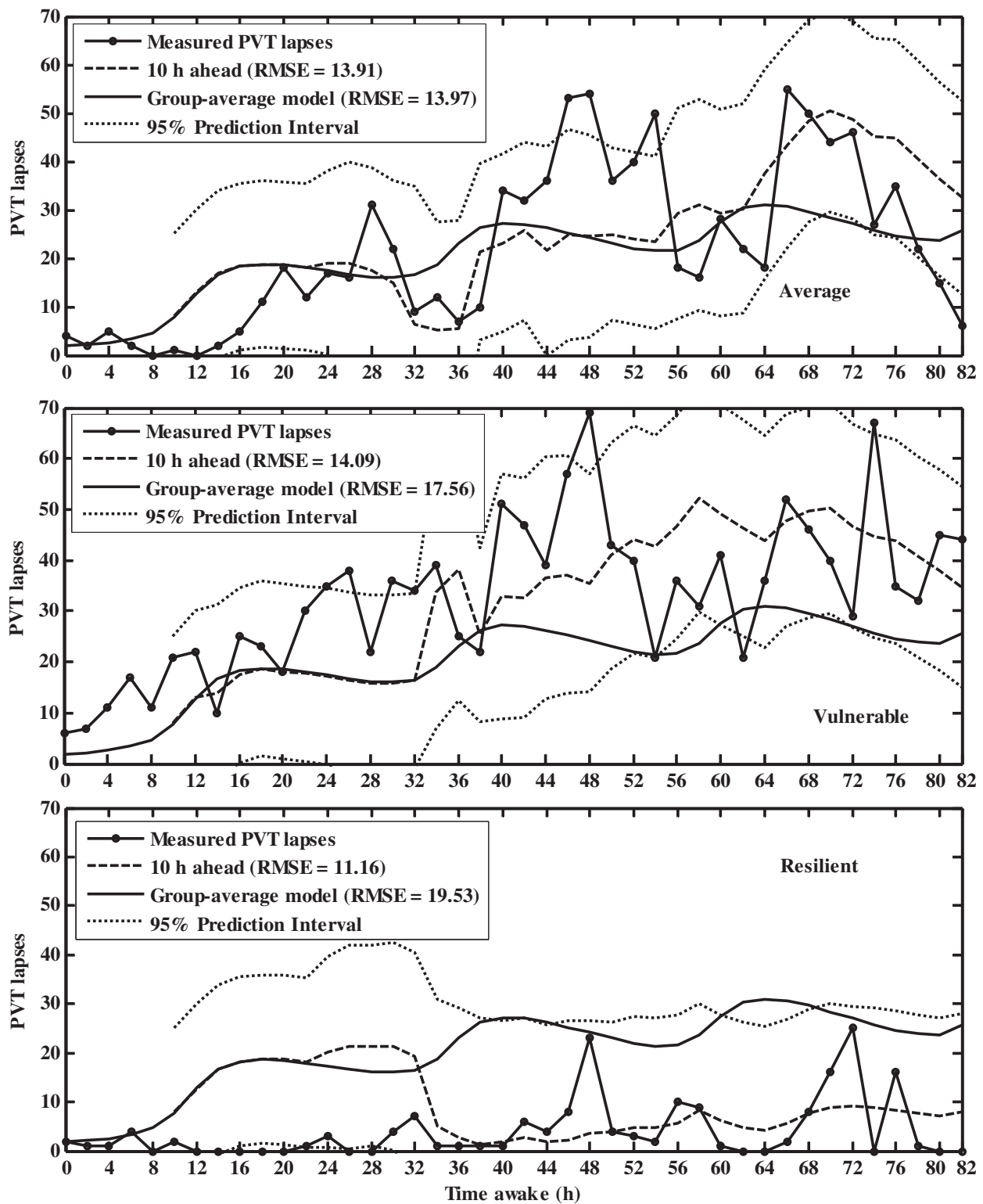
**Figure 7**—Individualized model predictions for three subjects with different behavioral phenotypes [average sensitivity to sleep loss (top), vulnerable to sleep loss (middle), and resilient to sleep loss (bottom)], with prior parameter values equal to those used in the group-average model and noise-level estimate $\hat{\sigma}^2$ set to 77.60. The solid circles in each of the panels represent the measured psychomotor vigilance task (PVT) lapses, measured every 2 hours. The dashed lines represent the 10-h-ahead predictions, whereas the dotted lines represent the corresponding analytically computed 95% prediction intervals. The solid lines in each panel represent the group-average model predictions.

loss, relatively average sensitivity to sleep loss, and relatively resilient to sleep loss. The three panels in Figure 7 show the observed PVT lapses (solid circles) for the average (top), vul-

nerable (middle), and resilient (bottom) subjects along with the group-average model predictions (solid lines), the 10-h-ahead predictions (dashed lines), and their corresponding 95%

PIs (dotted lines). In these calculations, we set the noise-level estimate $\hat{\sigma}^2 = 77.60$, as in Van Dongen.[10] For consistency, the RMSEs were computed between 10 and 82 h, and they indicated that the individualized predictions were up to 43% more accurate than the group-average model predictions for all three subjects, with the measurements falling almost completely within the corresponding analytically computed 95% PIs. We also compared the 10-h-ahead predictions with our previous method[11] and found that, on average, the new method reduced the prediction error by 10%.

## DISCUSSION

In this paper, we presented a new method based on the two-process model of sleep regulation that enabled individualized predictions of performance impairment represented by PVT lapses for subjects exposed to total sleep loss. The method advanced our previous work[11] in two important ways. First, it allowed model-parameter estimation and performance predictions as soon as the first performance measurement became available. This was achieved by combining *a priori* performance information with the individual's performance data through Bayesian inference. However, by retaining the strategy used in our previous work, where the nonlinear optimization problem of finding the best estimates of the two-process model parameters is transformed into a series of linear problems, the new method guaranteed unique estimates of the five parameters of the two-process model, avoiding brute-force, grid-search procedures.[10] Another important feature of the new method was that *a priori* performance information was optimally combined with the individual's performance data as a function of a user-provided estimate of the uncertainty (i.e., the noise level) in the measurements. As a result, the algorithm increased its trust in the *a priori* performance information as the estimate of the noise level in the measurements increased, and vice versa. Moreover, as more and more performance measurements were attained for an individual, the algorithm placed more and more trust in the measurements, de-emphasizing *a priori* performance information. The rate of increase of the trust placed in the measurements, however, became faster or slower as the user-provided uncertainty in the measurements decreased or increased, respectively.

Second, the current work, for the first time, provided statistically based error bounds around the model predictions in the form of prediction intervals. This was achieved by taking advantage of the linear representation of the two-process model proposed in our previous work,[11] which afforded the reformulation of the two-process model into an equivalent AR model and, thus, provided analytical expressions for estimating PIs about the model predictions, bypassing the need to first estimate confidence intervals of the model parameter estimates.[10]

Employing a simulated performance data set generated from the two-process model with selected parameters and with superimposed white Gaussian noise, we found that the parameters' estimates asymptotically converged to their true values as the number of performance data points increased and as the amount of noise in the performance data decreased, thereby confirming that the proposed method satisfied key convergence properties of adaptive algorithms.[18] Moreover, we observed that

performance predictions made over shorter horizons were more accurate than predictions made over longer horizons. This observation is because of the more frequent use of intermediate performance measurements by shorter horizons in updating the model-parameter estimates. However, intermediate measurements with low SNR could result in inaccurate models that, while yielding a good fit to the measurements, predict poorly.[18] We also analyzed the analytically computed 95% PIs and found, as expected, that their width was positively correlated with the amount of noise in the performance data. Moreover, as more performance measurements were available to the algorithm, the PIs converged to those empirically obtained through Monte Carlo trials, validating the accuracy of the PIs computed by the new method. Employing PVT data from a laboratory study, the proposed method yielded individualized predictions for three sleep-loss phenotypes that were up to 43% more accurate than group-average model predictions. Additionally, the corresponding 95% PIs almost entirely covered the entire set of measurements.

When comparing the results of the proposed method with those from our previous approach on the same simulated data, we observed average improvements in the accuracy of the model-parameter estimates of 70% and in the accuracy of the performance predictions of 80%. In particular, we found that the improvements in parameter estimation accuracy were more pronounced during the early phases of parameter estimation (typically, before obtaining 13 measurements) and when the SNR was low. This was because the algorithm places more trust on the *a priori* information during the early stages of adaptation (i.e., before the performance impairment trend could be completely learned) and when the estimated amount of noise in the data was large. Similarly, on the laboratory data set, the new method yielded a 10% average reduction in the prediction error.

We found the accuracy of the model-parameter estimates and of the performance predictions to be a function of the *a priori* selected values for the model parameters and the noise-level estimate of the performance measurements. We empirically observed that, as expected, the closer the *a priori* parameter values were to the *true* values underlying an individual, the more accurate were the predictions for that individual. Additionally, overestimation of the noise level in the measurements resulted in predictions biased toward the *a priori* information, whereas underestimation of the noise level did not yield necessarily worse predictions than the ones based on the correct value of the noise level. This was because the two-process model of sleep regulation is composed of sinusoidal and exponential components, which prevent a perfect fit of the model to non-smooth signals, such as additive white noise in the performance measurements. This precludes model overfitting and resulting deterioration of model predictions.

Despite the advances made by the new method for individualized performance prediction of sleep-deprived individuals, it does have limitations. In particular, this method suffers from the same limitations as any Bayesian approach, where a "good" choice of *a priori* parameter values and a "reasonable" estimates of the noise-level in performance data are key for obtaining optimal results.[18] Another limitation relates to the underlying assumption that measures of performance, such as PVT lapses, would be

available on a frequent basis. This may not be possible in some operational settings, since it may not be practical to discontinue work-related tasks for administering performance tests. Furthermore, we note that, although PVT lapse (selected as our predicted variable) is recognized as a widely used and sensitive measure to sleep loss and although PVT may be the most practical test for some operational environments,[37] it may not accurately reflect real, work-related performance of individuals.

Our future modeling efforts will focus on predicting performance under chronic sleep restriction conditions, where wake and sleep episodes alternate, precluding the availability of a large set of contiguous performance measurements from which individualized models can be obtained. This difficulty can be handled by our proposed method in that model adaptation would take place incrementally during wake episodes, and future performance levels would be predicted based on the latest parameter values. In addition, we will assess the performance of the proposed approach for the prediction of other outcome measures of performance, such as the Karolinska sleepiness scale,[38] the Stanford sleepiness scale,[39] the serial addition/subtraction task,[40] and the digit symbol substitution task.[41]

Considerable modeling efforts are still required to fully address the set of capability gaps identified at the Department of Defense-sponsored Fatigue and Performance Modeling Workshop.[1-3] However, the work described here provides significant contributions toward closing the research gaps and developing models that accurately predict cognitive performance impairments due to sleep deprivation at an individual level.

## 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 U.S. Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution.

## REFERENCE

1. Mallis MM, Mejdal S, Nguyen TT, Dinges DF. Summary of the key features of seven biomathematical models of human fatigue and performance. Aviat Space Environ Med 2004;75:A4-14.
2. Dinges DF. Critical research issues in development of biomathematical models of fatigue and performance. Aviat Space Environ Med 2004;75:A181-91.
3. Friedl KE, Mallis MM, Ahlers ST, Popkin SM, Larkin W. Research requirements for operational decision-making using models of fatigue and performance. Aviat Space Environ Med 2004;75:A192-9.
4. Webb WB, Levy CM. Effects of spaced and repeated total sleep deprivation. Ergonomics 1984;27:45-58.
5. Van Dongen HP, Baynard MD, Maislin G, Dinges DF. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep 2004;27:423-33.
6. Leproult R, Colecchia EF, Berardi AM, Stickgold R, Kosslyn SM, Van Cauter E. Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated. Am J Physiol Regul Integr Comp Physiol 2003;284:R280-90.
7. Wilkinson RT. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. J Exp Psychol 1961;62:263-71.
8. Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav Res Meth Instr Comp 1985;17:652-5.
9. Reifman J, Gander P. Commentary on the three-process model of alertness and broader modeling issues. Aviat Space Environ Med 2004;75:A84-8.
10. Van Dongen HP, 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.
11. Rajaraman S, Gribok AV, Wesensten NJ, Balkin TJ, Reifman J. Individualized performance prediction of sleep-deprived individuals with the two-process model. J Appl Physiol 2008;104:459-68.
12. Borbely AA. A two process model of sleep regulation. Hum Neurobiol 1982;1:195-204.
13. Daan S, Beersma DG, Borbely AA. Timing of human sleep: recovery process gated by a circadian pacemaker. Am J Physiol 1984;246:R161-83.
14. Viola AU, Archer SN, James LM, et al. PER3 polymorphism predicts sleep structure and waking performance. Curr Biol 2007;17:613-8.
15. Afari N, Buchwald D. Chronic fatigue syndrome: a review. Am J Psychiatry 2003;160:221-36.
16. Olofsen E, Dinges DF, Van Dongen HP. Nonlinear mixed-effects modeling: individualization and prediction. Aviat Space Environ Med 2004;75:A134-40.
17. Reifman J, Rajaraman S, Gribok AV. Moving towards individualized performance models. Sleep 2007;30:1081-2.
18. Cherkassky VS, Mulier F. Learning from data: concepts, theory, and methods. New York: Wiley, 1998.
19. Chatfield C. The analysis of time series: an introduction. 6th ed. Boca Raton: Chapman & Hall/CRC, 2004.
20. Indic P, Brown EN. Characterizing the amplitude dynamics of the human core-temperature circadian rhythm using a stochastic-dynamic model. J Theor Biol 2006;239:499-506.
21. Brown EN, Czeisler CA. The statistical analysis of circadian phase and amplitude in constant-routine core temperature data. J Biol Rhythms 1992;7:177-202.
22. Gribok AV, Buller MJ, Reifman J. Individualized short-term core temperature prediction in humans using biomathematical models. IEEE Trans Biomed Eng 2008;55:1477-87.
23. Achermann P, Borbely AA. Combining different models of sleep regulation. J Sleep Res 1992;1:144-7.
24. Borbely AA, Achermann P. Sleep homeostasis and models of sleep regulation. J Biol Rhythms 1999;14:557-68.
25. Klerman EB, Hilaire MS. On mathematical modeling of circadian rhythms, performance, and alertness. J Biol Rhythms 2007;22:91-102.

26. Achermann P, Borbely AA. Simulation of daytime vigilance by the additive interaction of a homeostatic and a circadian process. Biol Cybern 1994;71:115-21.
27. Ramsay JO, Silverman BW. Functional data analysis. 2nd ed. New York, Berlin: Springer, 2005.
28. Hansen PC. Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. Philadelphia: SIAM, 1998.
29. Marquardt DW. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 1970;12:591-612.
30. Kanaya F, Han TS. The asymptotics of posterior entropy and error probability for Bayesian estimation. IEEE Trans Inf Theory 1995;41:1988-92.
31. Reifman J. Alternative methods for modeling fatigue and performance. Aviat Space Environ Med 2004;75:A173-80.
32. Ord JK, Koehler AB, Snyder RD. Estimation and prediction for a class of dynamic nonlinear statistical models. J Am Stat Assoc 1997;92:1621-9.
33. Oleng NO, Gribok A, Reifman J. Error bounds for data-driven models of dynamical systems. Comput Biol Med 2007;37:670-9.
34. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. 3rd ed. Upper Saddle River, NJ: Prentice Hall, 1994.
35. Wesensten NJ, Killgore WD, Balkin TJ. Performance and alertness effects of caffeine, dextroamphetamine, and modafinil during sleep deprivation. J Sleep Res 2005;14:255-66.
36. Casella G, Berger RL. Statistical inference. 2nd ed. Australia; Pacific Grove, CA: Duxbury/Thomson Learning, 2002.
37. Balkin TJ, Bliese PD, Belenky G, et al. Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. J Sleep Res 2004;13:219-27.
38. Akerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. Int J Neurosci 1980;52:29-37.
39. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness - new approach. Psychophysiology 1973;10:431-6.
40. Thorne DR, Genser SG, Sing HC, Hegge FW. The Reed, Walter Performance Assessment Battery. Neurobehavioral Toxicology and Teratology 1985;7:415-8.
41. Wechsler D. The Psychometric Tradition - Developing the Wechsler Adult Intelligence Scale. Contemporary Educational Psychology 1981;6:82-5.
42. Bertsekas DP, Nedic A, Ozdaglar AE. Convex analysis and optimization. Belmont, MA: Athena Scientific, 2003.

## APPENDIX A

### Optimal Selection of the Parameter µ² in Eq. (9) [main text]

In this appendix, we describe the method employed to obtain an optimal value for the parameter $\mu^2$ in Eq. (9)

$$\arg\min_{0 \leq \gamma < 1}\left\{ \arg\min_{\overline{\mathbf{P}}}\left[ \left\|\mathbf{P} - \mathbf{y}\right\|_2^2 + \mu^2\left\|\widetilde{\mathbf{P}} - \widetilde{\mathbf{y}}\right\|_2^2 + \lambda^2\left\|\mathbf{L}_\gamma\overline{\mathbf{P}}\right\|_2^2 \right] \right\}, \quad \text{(A.1)}$$

which simultaneously accounts for the fit to the measurements $\mathbf{y}$ [the first term in Eq. (A.1)] and the fit to the prior performance data ỹ (the second term).

We obtained an optimal value for $\mu^2$ by minimizing the expression (i.e., the cost function) within the braces of the following equation

$$\arg\min_{\mu^2}\left\{ \left\|\mathbf{P}_{\mu^2} - \mathbf{y}\right\|_2^2 + tr[\text{cov}(\mathbf{P}_{\mu^2})] \right\}, \quad \text{(A.2)}$$

where $\mathbf{P}_{\mu^2}$ denotes the last $N$ elements of the solution of $\overline{\mathbf{P}}$ in Eq. (A.1), for a specific $\mu^2$ and for $\lambda^2$ set to an arbitrarily selected large number ($2^{1024}$). To avoid potential numerical problems with this procedure, we transformed Eq. (A.1) to its standard form,[28] solved it for a specific $\mu^2$ and for $\lambda^2$ set to an arbitrarily selected large number ($2^{1024}$), and transformed its solution back to the original problem [i.e., Eq. (A.1)]. The first term within the braces in Eq. (A.2) represents the fit of the solution of Eq. (A.1) to the performance data measurements $\mathbf{y}$, whereas the second term represents the trace of the covariance of $\mathbf{P}_{\mu^2}$, which is equivalent to the fit of the solution of $\overline{\mathbf{P}}$ to the *a priori*-generated performance data ỹ, i.e., the second term in Eq. (A.1). Noting that $\mathbf{P}_{\mu^2}$ and $tr[\text{cov}(\mathbf{P}_{\mu^2})]$ are functions of $\mu^2$ and $\lambda^2$, for which closed-form expressions exist,[28] and because Eq. (A.2) has a unique minimum (see below), the optimal $\mu^2$ can be obtained through standard numerical unconstrained optimization techniques, such as Levenberg-Marquardt or Gauss-Newton.

As $\mu^2$ increases, the first term in Eq. (A.2) increases monotonically while the second term decreases monotonically.[29] Therefore, the cost function in Eq. (A.2) has at most one minimum,[42] which corresponds to the optimal $\mu^2$. Because $tr[\text{cov}(\mathbf{P}_{\mu^2})]$ in Eq. (A.2) is proportional to the user-provided uncertainty $\hat{\sigma}^2$ in the measurements $\mathbf{y}$ (i.e., the noise level in $\mathbf{y}$), the optimal $\mu^2$ is a function of $\hat{\sigma}^2$. Thereby, as $\hat{\sigma}^2$ increases, the optimal $\mu^2$ increases, accentuating the trust in the prior performance data in Eq. (A.1), and vice versa.

## APPENDIX B

In this appendix, we show the equivalence between the two-process model of performance given by Eq. (5)

$$P(k) = \alpha - \alpha S(0)\gamma^{k-1} + \sum_{i=1}^{5}\overline{a}_i \sin\left[i\omega\left((k-1)T_s + \phi\right)\right], \quad \text{(B.1)}$$

and the 12th-order autoregressive (AR) model given by Eq. (10)

$$P(k) = \sum_{i=1}^{12} b_i P(k-i). \quad \text{(B.2)}$$

Note that for any set of values for the five parameters [$\alpha$, $\gamma$, $\beta$, $S(0)$, and $\varphi$], $P(k)$ in Eq. (B.1) is a solution to the 12th-order difference equation given by the expression within the brackets in Eq. (6)

$$\left[ (Z - \gamma)\left(\prod_{i=1}^{5}(Z^2 - 2\cos(i\omega T_s)Z + 1)\right)(Z - 1) \right]P(k) = 0, \quad \text{(B.3)}$$

where $Z$ denotes a shift operator, such that $Z^n\{P(k)\} = P(k+n)$. By expanding the linear, forward-difference operator in Eq. (B.3) into a 12th-degree polynomial in $Z$ and applying it to $P(k)$ in Eq. (B.1), we obtain the following difference equation

$$P(k+12) = c_1 P(k+11) + c_2 P(k+10) + .... + c_{12}P(k), \quad \text{(B.4)}$$

where $c_i$, with $i = 1,...,12$, are real numbers, which are fixed given the values of $\gamma$, $\omega$, and $T_s$. According to Eq. (B.4), we observe that $P(k+12)$ can be expressed as a linear combination of the previous 12 measurements or, alternatively, $P(k)$ can be expressed as a linear combination of $P(k-1)$, $P(k-2)$,..., and $P(k-12)$ as follows

$$P(k) = c_1 P(k-1) + c_2 P(k-2) + .... + c_{12} P(k-12)$$
$$= \sum_{i=1}^{12} c_i P(k-i). \quad (B.5)$$

By analyzing the correspondence between Eqs. (B.5) and (B.2), we concluded that the two-process model in Eq. (5) is equivalent to the 12th-order AR model given by Eq. (10), where $b_i = c_i$, with $i = 1,...,12$.