Alternative Methods for Modeling Fatigue and Performance

Jaques Reifman

Reifman J. *Alternative methods for modeling fatigue and performance.* **Aviat Space Environ Med 2004; 75(3, Suppl.):A173–80.**

The use of nonparametric approaches and semiparametric approaches for modeling fatigue and performance are analyzed. Nonparametric approaches in the form of stand-alone artificial neural networks and semiparametric (hybrid) approaches that combine neural networks with prior process knowledge are explored and compared with existing parametric approaches based on the two-process model of sleep regulation. Within the context of a military application, we explore two notional semiparametric approaches for real-time prediction of cognitive performance on the basis of individualized on-line measurements of physiologic variables. Initial analysis indicates that these alternative modeling approaches may address key technological gaps and advance fatigue and performance modeling. Most notably, these approaches seem amenable to predicting individual performance and quantitatively assessing the reliability of model predictions through estimation of statistical error bounds, which have eluded researchers for the last two decades.

Keywords: cognitive models, fatigue and performance models, hybrid models, gray models, artificial neural networks.

THE "FATIGUE AND Performance Modeling Workshop" brought together international experts who discussed the state-of-the-art of biomathematical models of fatigue, sleepiness, and performance (7). Assessment of the Workshop discussions and literature review of the presented models appear to indicate that, to varying extents, the basis of these mathematical models lie on the seminal two-process model of sleep regulation proposed by Borbély (3). The basic assumption is that sleep is regulated by two independent processes, a sleep-dependent homeostatic process (Process S) and a sleep-independent circadian process (Process C), which are summed together to estimate sleep propensity and the duration of sleep. Equally important, however, is the observation that the selected modeling approaches for implementing the two-process model are also essentially the same. Namely, they are based on a parametric modeling paradigm characterized by having a fixed structure where the model parameters are adjusted a priori to make the model fit a single data set and the model inputs (sleep/wake history and light exposure) are obtained either from prior records, when the models are employed to explain retrospective behavior, or estimated based on proposed scenarios, when the models are employed to forecast future behavior.

Therefore, given their similarity in genesis and modeling approaches, it is not surprising that no one model is systematically more accurate than the others (31), and it is unlikely that variations of the current approaches will address the existing gaps in neurobehavioral functional modeling, e.g., the ability to account for the effects of countermeasures, predict individual variability, and estimate model uncertainty (4,5,25).

In this paper, we make a first attempt to analyze, at a conceptual level, alternative approaches that could start addressing the existing gaps in fatigue and performance modeling, including the prediction of individual performance (as opposed to group-average performance) and the ability to quantitatively assess the reliability of model predictions through estimation of statistical error bounds. In particular, we explore the capabilities of nonparametric modeling approaches in the form of stand-alone artificial neural networks and semiparametric (hybrid) modeling approaches that combine neural networks and prior knowledge in the form of parametric models. The investigation is considered within the context of a military setting where it is assumed that physiologic variables and nonphysiologic variables, such as core body temperature, levels of light exposure, and sleep/wake history, are measured online for each subject for near real-time predictions of model parameters and estimates of individualized measures of fatigue and performance. Our intent is to expose these relatively recent modeling approaches to this community with the hope that they trigger further exploration and foster interdisciplinary collaboration, bringing new investigators and their modeling tools to advance the state-of-the-art in fatigue and performance modeling.

Alternative Modeling Approaches

Parametric approaches: The modeling approaches employed to date to predict measures of fatigue and performance fall into the class of "parametric" models. Parametric models are characterized by having a fixed structure derived from prior knowledge, which can take the form of existing empirical correlations, known mathematical equations, or fundamental first principles, such as the conservations of mass and energy. While widely applied, the parametric modeling ap-

From the Telemedicine and Advanced Research Center, U.S. Army Medical Research and Materiel Command, Ft. Detrick, MD.

Address reprint requests to: Jaques Reifman, Ph.D., who is a Senior Research Scientist at the U.S. Army Medical Research and Materiel Command, MCMR-AT 504 Scott St., Ft. Detrick, MD 21702-5012; reifman@tatrc.org.

proach requires comprehensive prior knowledge of the underlying phenomenon being represented in order to produce reliable and accurate predictions, where the required accuracy level depends on the application.

Because complete prior process knowledge is seldom available, parametric models are often augmented with other techniques, such as the well-known Kalman filter algorithm (9,15), for achieving improved state estimation and for estimating unmeasurable process parameters given the information contained both in the equations of the process dynamics and measurements made on the process. More recently, nonparametric approaches (10,16,17,26), i.e., purely data driven, structure-free methods, and semiparametric approaches that combine prior knowledge and data-driven methods (21,27,28,30,32,33), have been proposed as alternative, more flexible, and perhaps superior modeling paradigms than more traditional approaches. In what follows, we explore the characteristics of these two modeling paradigms for predicting measures of fatigue and performance.

Nonparametric approaches: Nonparametric approaches are characterized by a lack of prior model structure, where the model is synthesized without detailed knowledge of the underlying process and where the functional form of the model is conformed to the specifics of the particular process only after presentation of the data (28). Accordingly, nonparametric models are often termed "data-driven" models or "black-box" models, where the actual process data are used to derive the model. By contrast, parametric models are sometimes referred to as "white-box" models.

Artificial neural networks, in general, and multilayer feedforward neural networks (FNNs) (10,17,26), in particular, are one of the most popular types of nonparametric black-box modeling approaches. The popularity of FNNs stems from their underlying simplicity, power to approximate arbitrarily complex functions to any desired degree of accuracy (12), ability to model systems known only in terms of the system measurements when exact analytical equations are unavailable or difficult to develop, and ability to "learn" and achieve a desired overall behavior through the appropriate presentation of "training" input-output data pairs.

Learning (or training) is the process where the neural network approximates the function mapping from system inputs to outputs given a set of prior observations of its inputs and corresponding outputs (21). This is done by iteratively adjusting the network's internal free parameters, the so-called network weights w, typically in such a way as to minimize the prediction error ϵ defined as the square of the differences between the network predicted outputs and the desired (or observed) outputs. Training is often accomplished through the error back-propagation algorithm (26) or its variants (22) based on the computation of partial derivatives of ϵ with respect to the network weights w.*

Semiparametric approaches: Semiparametric modeling approaches (21,27,28,30,32,33) combine nonparametric models, such as neural networks, with fixed-form parametric models. Accordingly, they are often referred to as "gray-box" models or "hybrid" models. One basic idea of this approach is to embed prior knowledge about the process—in the form of parametric models—into the neural networks in order to impose internal structure so that different parts of the resulting model perform different tasks, allowing for clear interpretation of the response of the neural network models. This approach compartmentalizes the role of the neural networks into specific functions, reduces the network size and its training data requirements, and results in models with improved generalization and extrapolation than classical stand-alone black-box neural network models (21,28).

Yet, a more compelling rationale for developing hybrid models is to employ prior knowledge about the process to the maximum extent possible and complement the missing knowledge with information extracted from the process data. In this sense, the role of neural networks in hybrid models depends on the nature of the missing knowledge. For instance, in the case where a process behavior is well understood and is represented by first-principle balance equations, but the equations' parameters are time dependent, highly nonlinear, and difficult to infer, neural networks can be used to estimate the model parameters in real time on the basis of on-line process measurements. Conversely, when the process is not completely understood and the associated prior model is not exact, neural networks can be employed to account for the unmodeled process physics and compensate for the prior model inaccuracies by learning the residual differences between the prior model predictions and the observed (desired) process outputs.

Semiparametric (Hybrid) Architectures

Neural networks and prior knowledge can be combined in a number of different ways in the development of hybrid models. The three most predominant structural designs are illustrated in **Fig. 1**. In a serial semiparametric approach (Fig. 1a), the neural network estimates intermediate parameters, z, that are employed by the prior parametric model to describe the process behavior (21,30,33). This design is useful when the parameters z are unmeasurable, no a priori known model of the process parameters is available, or the parameters are highly nonlinear, being time- and state-space-dependent. In this design, the entire burden of estimating the process parameters is placed on the neural network model; however, the predictions z can be properly bounded and the prior model guarantees an appropriate output behavior.

Training (or determining the weights w) of the neural network in this hybrid design is more involving than training a stand-alone neural network because here the network outputs z are not measurable and input-output (x-z) data pairs are not available for training the network. Instead, the network is trained to learn the relationships between the inputs x and the hybrid outputs y. Accordingly, the network weights w are obtained by minimizing the prediction error ϵ, defined here as the

*For an in-depth discussion on artificial neural networks and other machine-learning algorithms, see Haykin (10) and Krogh et al. (17).

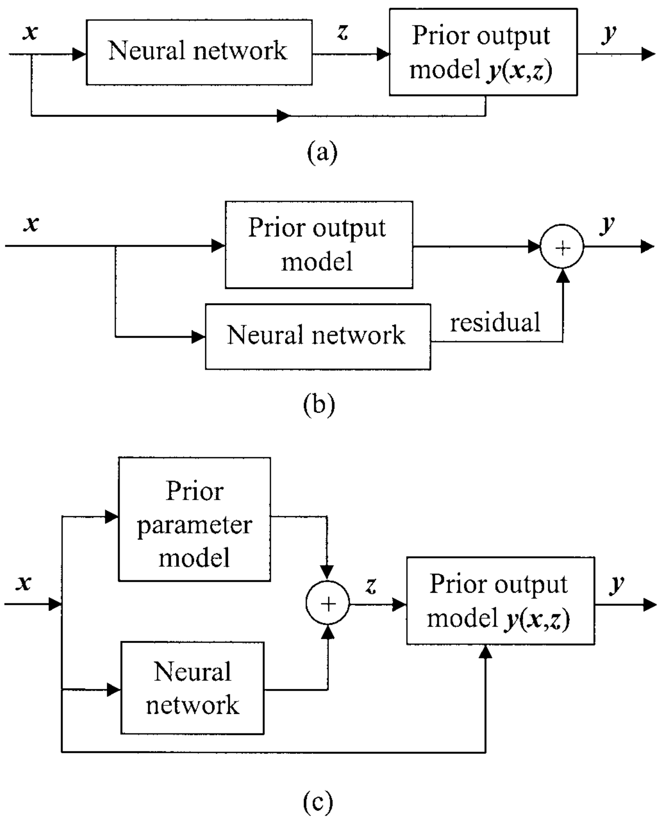

Fig. 1. Semiparametric (gray box or hybrid) approaches to combine prior knowledge with neural networks. a) Serial approaches use neural networks to estimate parameters of the prior model; b) Parallel approaches use the prior model as a guide to assist the neural network; c) Parallel/serial approaches use the neural network to account for inaccuracies of the prior parameter estimation model.

square of the differences between the predicted hybrid outputs $y(x,z)$ and the observed (or desired) process outputs representing a specific metric we wish to train the network to. In addition to the standard requirements of the back-propagation algorithm (26), determination of w involves the computation of the partial derivative of the hybrid outputs $y(x,z)$ with respect to the neural network's outputs z, whose computational complexity depends on the form of the prior model (21).

In a parallel semiparametric approach (Fig. 1b), the outputs of the neural network and the parametric model are combined to determined the total model output. The model serves as a first-order estimate of the process or a best guess at a process model and the neural network accounts for the unmodeled physics of the process behavior (27,32). Hence, this hybrid approach is useful when the process is not well understood and the derived parametric model is inexact. The output behavior of the hybrid cannot be guaranteed and the neural network may provide different contributions for different regions of the process input space. For regions of the space for which the parametric model is a good estimator, the neural network provides negligible contributions to the hybrid, and in regions where the model is inexact the network contributions are large. The neural network is trained to learn the residual function describing the differences between the parametric model predictions and the observed process outputs. Computing the network weights poses no additional requirements beyond those of training a stand-alone neural network.

In a parallel/serial approach (Fig. 1c), the outputs of the neural network and the prior parameter model are combined to estimate the model parameters z required as inputs to the output model. The prior parameter model serves as an idealized estimate of the output model parameters and the neural network compensates for its inaccuracies. In this configuration, prior knowledge improves performance by serving as a default estimate of the process parameters in the absence of data and as constraints (through the prior output model) that force the model to output predictions consistent with the physical process. Hence, this approach should provide considerable improvements when the system is used for extrapolation, as long as the prior parameter model is reasonably accurate over the range of extrapolation (28). The serial approach (Fig. 1a) represents a specific instance of this architecture in the case where the prior parameter model provides no contribution to the inference of the parameters z.

In a parallel/serial approach, the neural network learns the underlying relationships between the process inputs-outputs and the model parameters. Learning is obtained by adjusting the neural network weights w to minimize the error between the prior output model predictions and the observed process outputs. It should be easier to train such a network than the one in the serial approach, as here the neural network only needs to learn "part" of the process parameter as opposed to learning to predict the entire parameter. Because the network outputs are not measurable and a parametric model follows the parameter estimation step, like in the serial approach, the determination of w depends on the form of the parametric output model.

Comparison of Modeling Approaches

The characteristics selected in **Table I** to compare existing algebraic parametric approaches with nonparametric (neural network) approaches and semiparametric (neural network/prior knowledge) approaches for modeling fatigue and performance serve to illustrate the inherent features, capabilities, and limitations of each of the three modeling paradigms and to identify approaches that could address the shortcomings of the existing models (25).

Internal structure: A model is considered to have an internal structure if the various elements composing the model and their functions are clearly identified. One advantage of such models is that they can be easily interpreted and verified. Parametric models have a fixed internal structure derived from prior knowledge, generally in the form of existing empirical correlations, known mathematical equations, or first principles. For instance, the two-process model of sleep regulation (1) provides a good example of a parametric model, where the form and contribution of the two elements of the model, the homeostatic Process S and the circadian Process C, are clearly defined. Semiparametric models also have a defined internal structure with different parts of the model responsible for performing specific

TABLE I. COMPARISON OF EXISTING PARAMETRIC APPROACHES FOR MODELING FATIGUE AND PERFORMANCE WITH NONPARAMETRIC AND SEMIPARAMETRIC APPROACHES.

	Modeling Approaches		
Characteristics	Parametric (White Box)	Nonparametric (Black Box)	Semiparametric (Gray Box or Hybrid)
Internal Structure	Fixed structure derived from prior knowledge	Unknown structure based on data	Defined structure with different parts performing different tasks based on data and prior knowledge
Data Requirements	Modest	Large and data needs to be sampled at fixed intervals	Moderate, but data may need to be sampled at fixed intervals
Generalization	Limited to scenarios with similar sleeping patterns	Good interpolation but unreliable beyond limits of the training data	Better generalization and extrapolation than stand-alone black-box models
Prediction Reliability	Quantitative assessment of model prediction accuracy is not possible	Capability of providing statistically based confidence intervals and prediction intervals has been demonstrated	Capable of providing statistically based confidence intervals and prediction intervals
Model Inputs	Sleep/wake history and light exposure levels	Extended set of variables, such as sleep/wake history, light exposure, core body temperature, and electroencephalogram	Extended set of variables, such as sleep/wake history, light exposure, core body temperature, and electroencephalogram
Predicted Outputs	Need to be scaled to the appropriate objective or subjective measure of alertness and performance	Directly indicate desired measures of alertness and performance	Directly indicate desired measures of alertness and performance
Target Application	Group average performance based on off-line processing	Individual performance on the basis of on-line measurements and real-time processing, and group average performance based on off-line processing	Individual performance on the basis of on-line measurements and real-time processing

tasks. Nonparametric or "black-box" models, on the other hand, are structure free, where the unidentifiable functional form of the models is derived from the data.

Data requirements: Relatively modest amounts of data are required to estimate the parameters of the two-process model of sleep regulation. The decay and rise time constants used to determine the time course of Process S are derived from EEG slow-wave activity, whereas the shape and phase position of Process C are derived from physiologic variables such as core body temperature, or estimated indirectly from circadian sleep duration data (1). Conversely, stand-alone neural network approaches require large amounts of data, as their development (training) is solely based on observed process data. Hybrid approaches are also dependent on data, but less so than nonparametric approaches, because in hybrid systems, the neural network is only required to learn a portion of the process behavior. A more stringent requirement, however, is the frequency of the training data. In their development, both nonparametric and semiparametric approaches may require that the time-series data be equally spaced, i.e., sampled at fixed intervals with no missing values.

Generalization: A model generalization capability relates to the extent to which a model is able to correctly generalize its predictions to new, unanticipated conditions and scenarios within the limits of the developmental data. As reported by Van Dongen (31), the capability of the current generation of neurobehavioral functional models to generalize (and extrapolate) to unseen scenarios is somewhat limited, with one model performing better in one scenario and worse in another. Historically, whenever a parametric model failed to explain new data, new components were added to the model and model parameters were fitted to the new data in order to minimize the differences between "predicted" and observed values (8,14). Such an approach does not guarantee that the updated model will be able to generalize to the next unforeseen scenario.

The generalization capability of black-box neural network models is highly dependent on the amount of data available to train a network of a given size. If the available number of input-output training data pairs is not large enough relative to the number of adjustable network parameters w and the state space is not sampled sufficiently densely, the network generalization capabilities can be seriously compromised (23). For sufficiently large data sets, neural networks should perform arbitrarily well within the limits of the training data. However, black-box neural network models yield unreliable predictions when used beyond the limits of the training data, that is, when used for extrapolation. In this realm, hybrid approaches are expected to perform better than black-box neural network models, as generalization and extrapolation are confined only to the uncertain parts of the process while the basic (parametric) model is always consistent with prior knowledge and does not allow nonphysical variable interactions. Indeed, Psichogios and Ungar (21) have shown that—for the same number of training data points—the prediction error of hybrid systems is an order of magnitude lower than that of stand-alone neural network models, and that hybrid systems are capable of predicting the state of a process operating in a state-space regime that was not represented in the training data set while stand-alone neural networks failed. Thompson and Kramer report similar findings (28).

Prediction reliability: By their very nature, the current generation of fatigue and performance models does not possess the capability to quantitatively assess the precision of the model estimates for new data for which the outcomes are not known. Only the model by Moore-Ede et al. (20) attempts to infer lower and upper limits about the model predictions, but due to a lack of statistical underpinning in their approach, it is unlikely that the suggested limits around the predictions are meaningful (25).

Approaches based on the statistical bootstrap method (6) have been suggested to provide statistical error bounds about neural network predictions. First proposed by Tibshirani (29) to approximate confidence intervals of stand-alone neural network model predictions, the approach was subsequently extended by Heskes (11) to consider the more challenging and useful estimation of prediction intervals. The computationally intensive training of an ensemble of bootstrap neural network models is performed off-line, allowing for real-time and simultaneous prediction of the stand-alone neural network estimates alongside statistical confidence intervals and statistical prediction intervals. These statistical error bounds quantitatively determine within what bounds the predictions should be trusted for a predefined coverage probability, e.g., $\pm \epsilon$ with 95% confidence. Although to date its application has been limited to stand-alone neural networks, the bootstrap method could be directly extended to provide a statistically based, theoretically sound methodology for estimating the error bounds of hybrid model predictions.

Model inputs: The input requirements of the two-process model include: sleep/wake history or work hours and, in some cases, light exposure levels (19). As black-box neural network models and hybrid models are particularly suited for real-time applications, the input requirements could be considerably extended in these modeling paradigms to include on-line, real-time measures of an array of physiologic variables such as core body temperature, urine output, cortisol and melatonin levels, and EEG wave activity. Moreover, sleep/wake history could be monitored in real time via wearable wristband activity monitors (2), which could potentially reduce the uncertainty in this key input variable.

Predicted outputs: The outputs of the two-process model are internally scaled into arbitrary numerical units and subsequently mapped—using different methods, i.e., analogue scale, discrete scale—to different metrics corresponding to different subjective and objective measures of alertness and performance (25). However, such mapping is generally unknown and because each model output is mapped to its own arbitrary metric, the models cannot be directly compared with specific measurements of performance tests, such as the Psychomotor Vigilance Task (PVT) and the Karolinska Sleepiness Scale (KSS). In black-box neural network approaches and hybrid approaches, however, PVT, KSS, and any other measure of performance could be directly predicted by appropriately training the model to estimate each desired output, where a separate model would be trained for each desired performance metric.

Target application: Perhaps the Holy Grail in fatigue and performance modeling is the capability to predict individual variability. One approach for achieving this ultimate objective is to develop real-time models on the basis of individual, specific on-line data measurements. The on-line data measurements could be used to drive the model to respond to an individual's exposure to light and sleep/wake history, and to provide feedback to the model—by permitting adjustments of model parameters—on the basis of measurements of core body temperature and melatonin levels. While current parametric models do not lend themselves to such capabilities and are limited to predicting group-average performance (25), both stand-alone neural network models and hybrid models are inherently capable of achieving this objective. A notional framework along these lines focused on potential military applications is provided in the following section.

Hybrid Frameworks for Fatigue and Performance Modeling

The application of hybrid modeling approaches combining prior knowledge and neural networks within the context of future warfighter combat systems that will include improved electronics and personal computing equipment, communication systems, and enhanced clothing and biosensors that will be able to monitor the state of the soldier's health are considered in this section. The soldier's physiological status will be continuously monitored in real time on the basis of on-line measurements of physiologic and nonphysiologic variables that are fed to decision support systems to provide assistance in casualty prevention and management (13). We envision that core body temperature, heart rate, respiratory rate, hydration level and urine output, melatonin and cortisol levels, EEG wave activity, workload, sleep/wake history, light exposure levels, and perhaps even some measures of motivation and arousal could be captured by an array of noninvasive or minimally invasive biosensors to allow for near real-time assessment and management of a soldier's cognitive status.

A few different schemes for on-line monitoring and real-time predictions of measures of alertness and performance have been proposed in military settings.

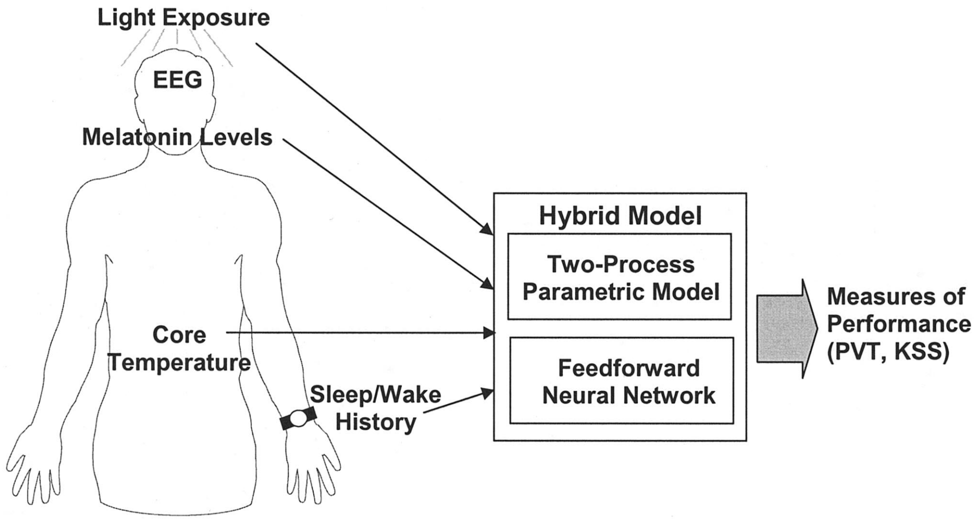

Fig. 2. Conceptual hybrid architecture combining artificial neural networks and the two-process model of sleep regulation. The hybrid receives as inputs on-line physiologic measurements and provides near real-time estimates of the subject's current cognitive performance level.

Belenky et al. (2) proposed a system for continuous monitoring of soldier performance where the system would not only measure physiologic variables and sleep/wake history (i.e., the process inputs x) but it would also directly measure the soldier's status (i.e., the process outputs y) through embedded reaction time tests. Makeig and Neri proposed an integrated dynamic system for managing shipboard work/rest scheduling (18). In their system, crew sleep/wake history and light exposure would be continuously monitored and objective measures of alertness would be inferred at discrete times on the basis of EEG spectral information and eye movement. The system conceptualized here is different from these prior schemes. Unlike the system of Belenky et al., here we only measure the system inputs and let the mathematical models make the inferences about the soldier's performance. And, unlike the system by Makeig and Neri, here the system makes continuous predictions, and these predictions are not based on the direct results of specific measurements, but rather on mathematical models of sleep regulation that take specific measurements as model inputs.

Within this context, we conceptualize the development of individual-specific hybrid models that provide near real-time predictions of soldier cognitive performance on the basis of individualized on-line data measurements. These models would be developed (trained) off-line based on the presentation of previously collected input-output data pairs to the hybrid. Once trained, the models would be used for near real-time predictions based on inputs provided from the on-line measurements. **Fig. 2** provides a high-level illustration of the approach. We consider the neural network to be represented by an FNN and the parametric model to be represented by a variant of the two-process model consisting of a sleep-dependent homeostatic component (Process S) and a sleep-independent circadian component (Process C). The hybrid takes the various measurements as inputs and provides objective or subjective measures of performance, such as PVT and KSS, as outputs, according to the specific output metric used for training. It is important to note that the two-process model is taken here as the parametric model only as an example to help illustrate the approach. The proposed hybrid architecture is general and in no way coupled to a specific parametric model.

In the serial approach (Fig. 1a), the FNN would dynamically compute unmeasurable process parameters z and provide them to the parametric two-process model to estimate measures of performance y. For example, the FNN could dynamically predict the shape and phase position of Process C from on-line measurements of light exposure, core body temperature, and melatonin levels while predicting the decay and rise time constants associated with Process S from EEG measurements. These predicted parameters along with measurements of sleep/wake history would be provided to drive the two-process model. Development of such a model requires that we first collect a large set of input-output data pairs, the inputs consisting of the physiologic and nonphysiologic measurements, and the outputs consisting of specific performance test results we desire the system to predict to train the neural network. Training consists of adjusting the FNN weights w so that the network provides the appropriate inferences z that force the differences between the hybrid predictions y and the observed performance test results (PVT, KSS) to be minimized. To estimate statistical confidence and prediction intervals about the hybrid inferences using the bootstrap method, we resample the training data with replacement to create N separate training data sets, train an ensemble of N hybrid systems with the N training sets, and apply the algorithms suggested by Heskes (11).

The serial approach is applicable if the model parameters are sensitive to variations in the measured variables, there are significant variations in the measured variables from subject to subject, and the parametric model is a good approximation of the underlying neurobehavioral function being modeled. The serial approach could also provide insight about the underlying model hypothesis. For example, inaccurate assumptions, such as which model parameters need to be dynamically updated, and missing modeled phenomena could be exposed from the lack of hybrid model fidelity. Also, by computing the partial derivative of the neural network outputs with respect to its inputs, insight could be gained about the relative importance of each measured physiologic variable in predicting the model parameters, leading to the selection of the most informative variables.

In the parallel approach (Fig. 1b), the FNN would capture the unmodeled sleep regulation behavior in the two-process model. For example, one of the basic assumptions of the two-process model is the independence of the homeostatic Process S and the circadian Process C, where the two components are independently calculated and simply added together (1). If this assumption were not strictly true and the two processes are indeed coupled, then the neural network would capture the unmodeled interactions between the two processes. In another example, we may consider the two-process model as providing a sort of "base-line model" for group predictions and the neural network as providing the necessary residual to account for individual variability and customize the predictions to specific subjects. The on-line measurements are provided to both the FNN and the two-process model and their combined contribution is the hybrid output. The two-process model would take the sleep/wake history as its input while the neural network could take any subset or all of the available on-line measurements, such as core body temperature, EEG wave activity, sleep/wake history, and light exposure levels. Analysis of the hybrid model prediction error as a function of variations of the neural network inputs could provide insight about the information content of each measured variable in predicting cognitive performance.

Training the FNN in this architecture also consists of minimizing ϵ, but in the parallel approach, the network learns the residuals that compensate for modeling uncertainties. Accordingly, this approach is useful when the contribution of the unmodeled physiology to the process behavior can be captured from the observable input-output measurements. Statistical error bounds about the hybrid predictions can be estimated with the bootstrap method (11).

As with any modeling paradigm, the hybrid approaches offer both advantages and disadvantages as alternative methods for modeling cognitive performance. The major potential advantages include: 1) on-line, near real-time estimation of cognitive performance; 2) prediction of individual performance; 3) models that are tuned to a specific performance metric, i.e., PVT, KSS, etc., without requiring mapping of the model's outputs; 4) models that generalize and extrapolate better than traditional parametric approaches and classical black-box neural networks; and 5) the ability to quantitatively assess the reliability of model predictions through the estimation of statistical error bounds. The most notable disadvantages include: 1) the large data requirements; and 2) the potential requisite that the time-series data be equally spaced with no missing data. The literature indicates that data sets ranging from 270 to 900 data points are necessary to develop each hybrid model (21,28). Obtaining such large data sets for each subject to develop individualized models could prove impractical until the warfighter physiological status monitoring system is fielded, allowing for routine data collection (13). The lack of missing performance test data, in particular during sleep periods, however, could be circumvented by assuming perfect model predictions during these time intervals, that is, by assuming a zero contribution of these data points to the overall prediction error ϵ.

Conclusions

In this paper, we compared and contrasted the capabilities of the parametric two-process model of sleep regulation with nonparametric and semiparametric (hybrid) modeling approaches. We find the attributes of hybrid models to be particularly attractive within the context of military relevant settings where soldier cognitive status could be estimated in near real time on the basis of on-line measurements of physiologic and nonphysiologic variables. Initial analysis indicates that, within this context, hybrid models could potentially address key limitations of current state-of-the-art approaches, such as the capability to predict individual performance and quantitatively assess the reliability of model predictions through estimation of statistical error bounds about the model predictions. Two conceptual hybrid models that combine neural networks with prior knowledge in the form of the two-process model of sleep regulation are proposed to illustrate the approach, but the methodology is generic and applicable to any number of parametric models.

Modeling and simulation cut across all domains of science. Hence, it would behoove the fatigue and performance modeling community to reach out and explore alternative modeling techniques and algorithms that have been suggested and successfully implemented in other scientific domains. For example, in addition to the approaches proposed herein, stochastic techniques in the form of the maximum likelihood, the Kalman filter, and the Luenberger observer should be explored as alternative means to improve process state and process parameter estimation. This need to reach out and borrow ideas and techniques from other domains of science is particularly critical if we aspire to address the existing technological gaps in fatigue and performance modeling which have eluded researchers for the past two decades. Oftentimes, what seems undoable using one methodology becomes quite doable when the problem is analyzed through a different angle, leading to nonlinear advances. For example, we may not need to master group-average model predictions before exploring means of predicting individual performance. Due to large person-to-person variability in cognitive performance, it may turn out to be easier to develop reliable and accurate models of individualized measures of performance on the basis of an individual's on-line physiologic measurements than group-average performance models.

ACKNOWLEDGMENTS

The author is thankful to the reviewers and Harris Lieberman for providing helpful feedback. The author was supported in part by the Combat Casualty Care and the Military Operational Medicine research programs of the U.S. Army Medical Research and Materiel Command, Ft. Detrick, MD.

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

REFERENCES

1. Acherman P, Borbély AA. Simulation of daytime vigilance by the additive interaction of a homeostatic and a circadian process. Biol Cybern 1994; 71:115–21.
2. Belenky G, Balkin TJ, Redmond DP, et al. Sustaining performance during continuous operations: the U.S. Army's sleep management system. In: Friedl K, Lieberman H, Ryan DH, Bray GA, eds. Countermeasures for battlefield stressors: Pennington center nutrition series. Baton Rouge, LA: Louisiana State University Press; 1999:197–205.
3. Borbély AA. A two process model of sleep regulation. Hum Neurobiol 1982; 1:195–204.
4. Dinges DF, Achermann P. Future considerations for models of human neurobehavioral function [Commentary]. J Biol Rhythms 1999; 14:598–601.
5. Dinges DF. Critical research issues in development of biomathematical models of fatigue and performance. Aviat Space Environ Med 2004; 75 (3, Suppl.):A181–91.
6. Efrom B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
7. Fatigue and performance modeling workshop. 2002; Retrieved 10 June 2003 from http://fatigue.anteon.com/.
8. Folkard S, Åkerstedt T, Macdonald I, et al. Beyond the three-process model of alertness: Estimating phase, time on shift and successive night effects. J Biol Rhythms 1999; 14:577–87.
9. Hassberger JA, Lee JC. A simulation-based expert system for nuclear power plant diagnostics. Nucl Sci Eng 1989; 102:153–71.
10. Haykin S. Neural networks: A comprehensive foundation. 2nd ed. Englewood Cliffs, NJ: Prentice Hall; 1999.
11. Heskes T. Practical confidence and prediction intervals. In: Mozer MC, Jordan MI, Petsche T, eds. Advances in neural network information processing systems 9. Cambridge, MA: The MIT Press; 1997; 176–82.
12. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw 1989; 2:359–66.
13. Hoyt RW, Reifman J, Coster TS, Buller MJ. Combat medical informatics: Present and future. Proceedings of the AMIA 2002 Annual Symposium; 2002 Nov 9–13; San Antonio, TX. Bethesda, MD: American Medical Informatics Association; 2002: 335–9.
14. Johnson MJ, Belenky G, Redmond DP, et al. Modulating the homeostatic process to predict performance during chronic sleep restriction. Aviat Space Environ Med 2004; 75 (3, Suppl.): A141–6.
15. Kalman RE, Bucey KS. New results in linear filtering and predictive theory. Transactions of the ASME. J Basic Eng 1965; 83D: 95–108.
16. Kramer MA. Autoassociative neural networks. Comput Chem Eng 1992; 16:313–28.
17. Krogh A, Palmer RG, Hertz JA. Introduction to the theory of neural computation. New York: Perseus Books Group; 1991.
18. Makeig S, Neri DF. A proposal for integrated shipboard alertness management. In: Friedl K, Lieberman H, Ryan DH, Bray GA, eds. Countermeasures for battlefield stressors: Pennington center nutrition series. Baton Rouge: Louisiana State University Press; 1999:1–8.
19. 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 (3, Suppl.):A4–14.
20. Moore-Ede M, Heitmann A, Croke D, et al. Circadian alertness simulator for fatigue risk assessment in transportation: Application to reduce frequency and severity of truck accidents. Aviat Space Environ Med 2004; 75 (3, Suppl.):A107–18.
21. Psichogios DC, Ungar LH. A hybrid neural network: First principles approach to process modeling. Am Inst Chem Eng J 1992; 38:1499–511.
22. Reifman J, Vitela JE. Accelerating learning of neural networks with conjugate gradients for nuclear power plant applications. Nucl Technology 1994; 106:225–41.
23. Reifman J, Feldman EE, Wei TYC, Glickert RW. An intelligent emissions controller for fuel lean gas reburn in coal-fired power plants. J Air Waste Manag Assoc 2000; 50:240–51.
24. Reifman J, Feldman EE. Multilayer perceptron for nonlinear programming. Comput Ops Res 2002; 29:1237–50.
25. Reifman J, Gander P. Commentary on the three process model of alertness and broader modeling issues. Aviat Space Environ Med 2004; 75 (3, Suppl.):A84–8.
26. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representation by error propagation. In: Rumelhart DE, McClelland JC, eds. Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge, MA: The MIT Press; 1986:318–62.
27. Stubberud SC, Lobbia RN, Owen M. Adaptive state estimation using artificial neural networks. In: Dagli CH, ed. Intelligent engineering systems through artificial neural networks. New York: ASME Press; 1995:5:531–6.
28. Thompson ML, Kramer MA. Modeling chemical processes using prior knowledge and neural networks. Am Inst Chem Eng J 1994; 40:1328–40.
29. Tibshirani R. A comparison of some error estimates for neural network models. Neural Comput 1996; 8:152–63.
30. van Can HJL, Hellinga C, Luyben KCAM, et al. Strategy for dynamic process modeling based on neural networks in macroscopic balances. Am Inst Chem Eng J 1996; 42:3403–18.
31. Van Dongen HPA. Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviat Space Environ Med 2004; 75 (3, Suppl.):A15–36.
32. Vilim RB, Garcia HE, Chen FW. An identification scheme combining first principle knowledge, neural networks, and the likelihood function. IEEE Trans Control Sys Tech 2001; 9:186–99.
33. Wilson JA, Zorzetto FM. A generalized approach to process state estimation using hybrid artificial neural networks/mechanistic models. Comput Chem Eng 1997; 21:951–63.