

# A Real-Time Algorithm for Predicting Core Temperature in Humans

Andrei V. Gribok, Mark J. Buller, Reed W. Hoyt, and Jaques Reifman

**Abstract**—In this paper, we present a real-time implementation of a previously developed offline algorithm for predicting core temperature in humans. The real-time algorithm uses a zero-phase Butterworth digital filter to smooth the data and an autoregressive (AR) model to predict core temperature. The performance of the algorithm is assessed in terms of its prediction accuracy, quantified by the root mean squared error (RMSE), and in terms of prediction uncertainty, quantified by statistically based prediction intervals (PIs). To evaluate the performance of the algorithm, we simulated real-time implementation using core-temperature data collected during two different field studies, involving ten different individuals. One of the studies includes a case of heat illness suffered by one of the participants. The results indicate that although the real-time predictions yielded RMSEs that are larger than those of the offline algorithm, the real-time algorithm does produce sufficiently accurate predictions for practically meaningful prediction horizons ( $\sim 20$  min). The algorithm reached alert ( $39^\circ\text{C}$ ) and alarm ( $39.5^\circ\text{C}$ ) thresholds for the heat-ill individual but did not even attain the alert threshold for the other individuals, demonstrating the algorithm's good sensitivity and specificity. The PIs reflected, in an intuitively expected manner, the uncertainty associated with real-time forecast as a function of prediction horizon and core-temperature variability. The results also corroborate the feasibility of “universal” AR models, where an offline-developed model based on one individual's data could be used to predict any other individual in real time. We conclude that the real-time implementation of the algorithm confirms the attributes observed in the offline version and, hence, could be considered as a warning tool for impending heat illnesses.

**Index Terms**—Autoregressive (AR) models, core-temperature predictions, real-time prediction.

Manuscript received March 2, 2009; revised December 23, 2009; accepted February 11, 2010. Date of publication April 5, 2010; date of current version July 9, 2010. This work was supported in part by the Military Operational Medicine Research Area Directorate of the U.S. Army Medical Research and Materiel Command.

A. V. Gribok was with the Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Frederick, MD 21702 USA. He is now with the Department of Nuclear Engineering, The University of Tennessee, Knoxville, TN 37996 USA.

M. J. Buller is with Brown University, Providence, RI 02912 USA, and also with the U.S. Army Research Institute of Environmental Medicine, Natick, MA 01760 USA.

R. W. Hoyt is with the U.S. Army Research Institute of Environmental Medicine, Natick, MA 01760 USA.

J. Reifman is with the Department of the Army, Department of Defense Biotechnology High Performance Computing Software Applications Institute for Force Health Protection, and Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Frederick, MD 21702 USA (e-mail: jaques.reifman@us.army.mil).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITB.2010.2043956

## I. INTRODUCTION

ALTHOUGH heat illnesses are presumably preventable, they are difficult to predict because in some circumstances, such as at the height of a military operation or during an athletic competition, humans may ignore early warning signs of a rising core temperature and impending heat illnesses [1]. Different heat strain indexes have been proposed to evaluate an individual's susceptibility to heat stress [2], [3]; however, they lack predictive capabilities and can only evaluate the current physiological state of the individual when it is already too late for proactive response.

In our previous work [4], [5], we showed that, due to the large thermal inertia of the human body, the core temperature in humans can be accurately predicted with an autoregressive integrated (ARI) model for up to 20 min ahead of time. The 20-min-ahead prediction horizon is long enough to be practically useful in an early warning system that could be worn by athletes and soldiers to forecast rising core temperatures during intense physical activity in hot-weather conditions. However, this relatively long prediction horizon was obtained using core-temperature data that had been smoothed offline by a global filtering technique, which requires the availability of the entire time-series dataset. Obviously, such requirement is not met in real-time applications, where future core-temperature data are unknown and only current and previous values are available. In addition, the global filtering technique relies on regularized differentiation of the core-temperature data, which is computationally intensive and may be beyond the computational power of wearable devices.

In this paper, we describe an algorithm for predicting core temperature in real time and investigate its performance against the offline version. We extended our earlier work to include a real-time filtering technique that is less computationally demanding and an AR model that, unlike ARI models, does not require differentiation. The main reason for applying ARI models for time-series predictions is to handle nonstationarity arising from variations of the signal's statistics. Although the offline version of the algorithm uses first-order differentiation of the core-temperature signal to ensure mean-value stationarity, after additional investigations, we found that the low-order statistics of the core-temperature signal, such as the mean value and the autocorrelation function, exhibited very mild variations with respect to time. Most likely, this is because the core-temperature signal is a physiologically very tightly regulated signal, with low-order statistics that remain practically unchanged through time. Comparison of ARI and AR models for real-time predictions showed no additional benefits for using the more complicated ARI model. In this

study, we used AR models for both offline and real-time predictions.

In addition to point predictions, we also estimated prediction intervals (PIs) based on a previously developed algorithm [5] to provide a measure of reliability of the point predictions in real time. Several tests with field-study data involving military activities, including a case in which one subject's core temperature reached a critical threshold, indicate that the algorithm's real-time performance degrades in comparison with its offline version. However, the algorithm is still a valuable tool, providing real-time point predictions and associated PIs that could be used as an early warning of impending heat illnesses.

## II. METHODS

The real-time core-temperature prediction algorithm consists of two main components: data filtering and predictive model. The data filtering component was implemented using a Butterworth zero-phase, low-pass filter of order five with a cutoff frequency of  $4.25 \times 10^{-4}$  Hz. The cutoff frequency was selected based on the analysis of the power spectrum of the core-temperature signal such that approximately 99% of the signal's variance was contained in the range below the cutoff frequency. The Butterworth filter was selected instead of other alternatives because it has the flattest response in the pass band, thus, producing smooth signals that can match the smoothness of the regularized signals in the offline version of the algorithm. To eliminate the phase shift between the raw and the filtered signals introduced by the Butterworth filter, we applied a forward-backward filtering technique [6] in which the raw signal was first filtered forward in time and then the same filter was applied backwards to the forward-filtered signal.

The low-pass filter is often coupled with data downsampling to reduce the Nyquist frequency of the signal. The sampling frequency of the core-temperature signal, recorded by a telemetry core-temperature pill, is one sample per minute. This sampling frequency is rather high because the metabolic processes governing the changes in core temperature occur at much longer time scales. To remove the high-frequency noise introduced by the short-sampling interval, the core-temperature signal was first downsampled to 5-min intervals by keeping only every fifth-sampled signal, before applying the low-pass filter.

A Butterworth filter uses previously filtered and raw (i.e., unfiltered) signals to produce a filtered signal  $\tilde{y}_t$  at time  $t$

$$\tilde{y}_t = \sum_{i=0}^n \theta_i y_{t-i} - \sum_{j=1}^n \varphi_j \tilde{y}_{t-j} \quad (1)$$

where  $y_t$  denotes the raw signal at time  $t$ ,  $n$  denotes the order of the filter, and  $\theta$  and  $\varphi$  represent the vectors of filter coefficients. As each new core-temperature sample  $y_t$  became available at time  $t$ , it was incorporated into the vector of core-temperature samples, and the filter was iteratively applied to the entire time series, first forward, from the first sample to the last sample at time  $t$ , and then backward to the first sample. The forward-backward filtering requires the availability of an initial batch of signal samples because it uses the information of the flanking samples to compensate for the phase shift. For real-time appli-

cations, this means that there will be an initial "waiting" period, with length equal to the order of the model, before the filter and the predictive model can be applied. During this waiting period, we used raw, unfiltered core-temperature values to represent  $\tilde{y}_{t-j}$  in (1). We also assumed that the time required to filter the data was negligible when compared with the sampling interval.

We used an AR model of order  $m$  to make near future core-temperature predictions. Given filtered signals  $\tilde{y}_{t-i}$ ,  $i = 0, \dots, m-1$ , an AR model produces an output or predicted signal  $\hat{\tilde{y}}_{t+1}$ , at time  $t+1$ , through a linear combination of the antecedent, filtered core-temperature samples

$$\hat{\tilde{y}}_{t+1} = \sum_{i=1}^m b_i \tilde{y}_{t-i+1} + \varepsilon_{t+1} \quad (2)$$

where  $b$  denotes the vector of  $m$  unknown AR coefficients and  $\varepsilon_{t+1}$  represents white noise with unknown variance. To make predictions  $M$  time steps ahead, we iteratively used (2)  $M$  times, substituting the unobserved signals at  $t \geq (t+1)$  in the summation by their corresponding predicted values. As discussed earlier, the order of the model  $m$  specifies the required initial waiting period for which data samples need to be collected before real-time predictions can be made.

Before applying the AR model, we must first determine the AR coefficients  $b$  using some "training" data. In our earlier work [4], we identified a remarkable property of ARI models for core-temperature predictions in which models trained for one individual can be applied to predict the temperature of other individuals provided the individuals have similar anthropomorphic characteristics and the training data are representative of the range of activities under which the model is applied. Accordingly, these so-called "universal" or "portable" models offer the possibility of using data from only one individual to train an AR model offline to determine the coefficients  $b$ , and subsequently applying the model to predict all other individuals in real time. To smooth the training data employed to develop the universal model, we applied the same Butterworth filter as the one subsequently used in real-time predictions for the entire training data, resulting in core-temperature samples that were more correlated with each other than the samples of the original raw signals. Although samples that are more correlated are easier to predict, the higher correlation also causes the design matrix of the least squares (LS) method used to determine the coefficients  $b$  to become ill-conditioned, yielding models with large variance. To alleviate this problem and obtain consistent AR coefficients, we extended the LS method by adding a penalty function and solving a regularized LS problem [4], [7].

The inclusion of a penalty function in the solution of a regularized LS problem has implications in the selection of the order  $m$  of the AR model. Because regularization constrains the values of the coefficients  $b_i$ ,  $i = 1, 2, \dots, m$ , well-established criteria for selecting the order of an AR model, such as the Akaike information criterion [8], are not applicable, as the fitting error may not decrease with the increasing model order. This fact allows the use of regularized models of higher order without overfitting concerns. The advantage of using a higher order model is that it accommodates higher frequencies present in the signal,

thus reducing prediction lag. Here, we tried models of different orders, however, the lower-order models inevitably introduced more lag in the predicted signal. The selected model of order  $m = 25$  represents an empirically obtained compromise between a reduction in prediction lag and a minimization of the waiting period.

A fundamental difference between offline and real-time predictions lies in the way the testing data are filtered. For the offline predictions in our earlier work [4], the entire testing data are filtered using regularized differentiation of the core-temperature signal. For real-time predictions, the Butterworth filter can only use the testing data up to the current time, making it difficult for the filter to accurately smooth the most recent data samples, which carry the majority of the predictive information. Such filtering inaccuracy leads to less consistent and delayed predictions.

The application of the AR model for predictions was identical for both offline and real-time implementations. For all simulations, the prediction horizon, unless otherwise noted, was set to 20 min, and the prediction accuracy was evaluated using the root mean squared error (RMSE), defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - \tilde{y}_i)^2}. \quad (3)$$

We chose to compute the RMSE between the predicted  $\hat{y}$  and the filtered signal  $\tilde{y}$ , as opposed to the raw and unfiltered signal  $y$ , because  $y$  was laden with noise and outlier values [4], which would have yielded artificially large RMSEs.

In many safety-critical applications, providing single-point predictions may not be sufficient. A measure of the reliability of the point predictions may be required to assess the uncertainty of the predicted values. In earlier work [5], we developed a technique based on the statistical bootstrap method [9] to estimate prediction reliability in the form of PIs. The technique relies on the idea of model resampling [5] rather than data resampling [9], where a population of models is built based on blocks of data that are randomly drawn from the original time series to form an empirical distribution of models. Models are resampled from the distribution to make predictions and construct a distribution of model predictions from which the PIs are inferred. It should be stressed that PIs are different from traditional confidence intervals used in statistics, since they account for two types of uncertainty: in the model and in the data. We used the aforementioned technique to estimate the PIs for the core-temperature point predictions. Accordingly, the PI for  $\hat{y}_{t+1}$  in (2) can be estimated as [5]:

$$\text{PI} = \hat{y}_{t+1} \pm Z_{\alpha/2} \sigma(\text{pred}) \quad (4)$$

where  $Z_{\alpha/2}$  denotes the prediction factor associated with an  $\alpha\%$  type I error and  $\sigma(\text{pred})$  denotes the standard deviation of the prediction error. Here, we set  $Z_{\alpha/2} = 2.98$  [5].

To demonstrate the performance of the real-time algorithm, we used data from two field studies involving a total of ten subjects performing military-related field exercises. Both studies were approved by the Institutional Review Board of the U.S. Army Research Institute of Environmental Medicine, Natick,

MA and the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD. In both studies, the core-temperature data were measured using radio-thermometer analog pills (HQ Inc., Palmetto, FL) and retrieved *post hoc*. The pills were ingested each evening  $\sim 8$  h prior to the data collection and had the following technical characteristics: size: 22.4-mm length and 10.9-mm diameter; frequency: 262 kHz; temperature range: 30°C–45°C, with accuracy of  $\pm 0.1$ °C; transmission method: near-field magnetic link; and sampling rate: 10 s to hourly. The core temperature is considered to be an accurate reflection of the thermal state of an individual, although a very recent study [10] suggests that the accuracy may be dependent on time of day.

The first study consists of core-temperature data collected every minute from eight U.S. Marine Corporations volunteers [age: 25 year (SD 3.2); height: 174 cm (SD 6.7); weight: 71.6 kg (SD 7.9); body fat percentage: 15.9% (SD 7.1), mean and standard deviation (SD)] during a four-day field exercise. Each 10-h day involved a 3-mi morning march to a shooting range, followed by day-long exercises and rotations within firing stations, and a march back via the same route in the evening. Subjects wore air-permeable battle dress uniform [thermal resistance = 1.32 m<sup>2</sup>·(K/W)] and, when marching, carried a pack load of 26  $\pm$  1.0 kg. The ground temperature during the day was 29.8°C (SD 0.5), and the dew point and wind speed were 21.1°C (SD 0.5) and 4.2 m/s (SD 0.5), respectively.

The second study consists of core-temperature data recorded every minute for  $\sim 8$  h from two subjects, a cadet [age: 21 year; height: 175.0 cm; weight: 73.9 kg; body fat percentage: 17.9] and a soldier [age: 22 year; height: 170.0 cm; weight: 68.0 kg; body fat percentage: 13.3], involved in war games. The soldier and the cadet carried loads of 35 and 45 kg, respectively, and wore utility uniforms. The ground temperature during the day was 33.0°C (SD 0.5), and the relative humidity and wind speed were 70.0% (SD 1.0) and 1.0 m/s (SD 0.5), respectively. The cadet's core temperature underwent a sudden increase starting at 12:20 h and reached an extreme value of 39.5°C around 12:50 h. Although the elevated value of his core temperature was unknown at that time, while passing through a monitoring station the cadet was immediately pulled from the exercise because he exhibited visible signs of heat exhaustion. This dataset is particularly valuable because it presents an opportunity to test the point predictions and PIs estimates at difficult-to-obtain, extreme temperature conditions. The time-series data from the two studies were downsampled to 5-min intervals before applying the filtering and prediction algorithms.

### III. RESULTS AND DISCUSSION

To compare and contrast the performance of the offline and real-time versions of the algorithms, we performed three sets of simulations involving the two field studies. Each simulation was performed twice: once mimicking the real-time predictions and the other the offline predictions. In the first simulation, we employed data from the first study to investigate the performance of the algorithm when one portion of a subject's data was used to train the model, i.e., to obtain the coefficients  $b$  of the AR

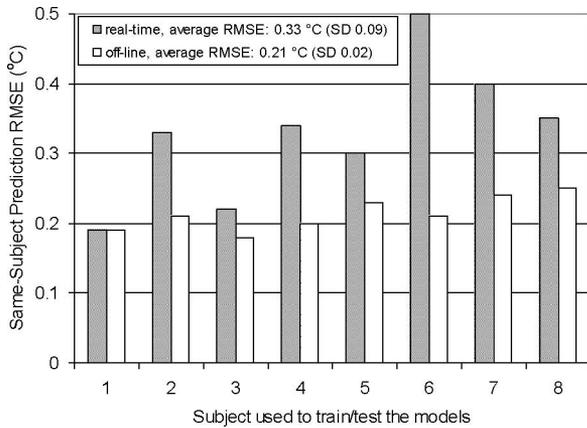


Fig. 1. Real-time and offline root mean squared errors (RMSEs) for the same-subject predictions, for each of the eight subjects in the first field study.

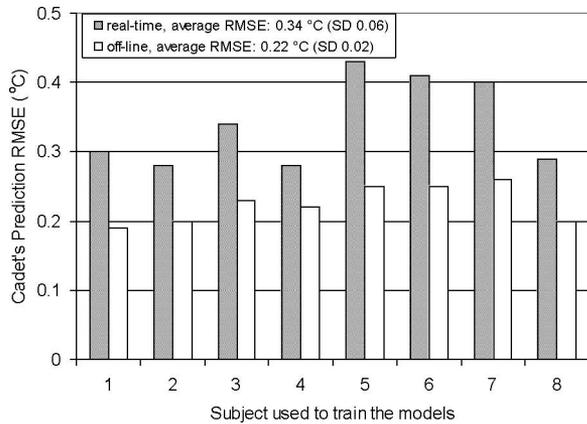


Fig. 2. Real-time and offline RMSEs for the cross-subject and cross-study 20-min-ahead predictions of the cadet's core temperature, using the models from each of the eight subjects of the first study.

model, and another portion of the same subject's data was used to test (or assess) the model's predictions. Specifically, for each of the eight individuals in the first study, we identified the two days with the longest available core-temperature records and used one of those days to train the model and the other to test the model. Fig. 1 shows the testing data RMSEs for such same-subject predictions. The average RMSE over the eight subjects for the real-time predictions was  $0.33\text{ }^{\circ}\text{C}$  (SD 0.09) and for the offline predictions it was  $0.21\text{ }^{\circ}\text{C}$  (SD 0.02).

In the second simulation, we investigated the performance of the algorithms through a cross-subject and cross-study test, involving the both field studies. We used the models developed for each of the eight subjects from the first simulation described earlier to predict the core-temperature profile for each of the two subjects, the cadet and the soldier, of the second study. Figs. 2 and 3 show the prediction RMSEs for the cadet and the soldier, respectively. For the cadet's predictions, the average RMSE was  $0.34\text{ }^{\circ}\text{C}$  (SD 0.06) for the real-time algorithm and  $0.22\text{ }^{\circ}\text{C}$  (SD 0.02) for the offline algorithm. For the soldier, the average RMSE was  $0.22\text{ }^{\circ}\text{C}$  (SD 0.02) for the real-time algorithm and  $0.17\text{ }^{\circ}\text{C}$  (SD 0.05) for the offline algorithm.

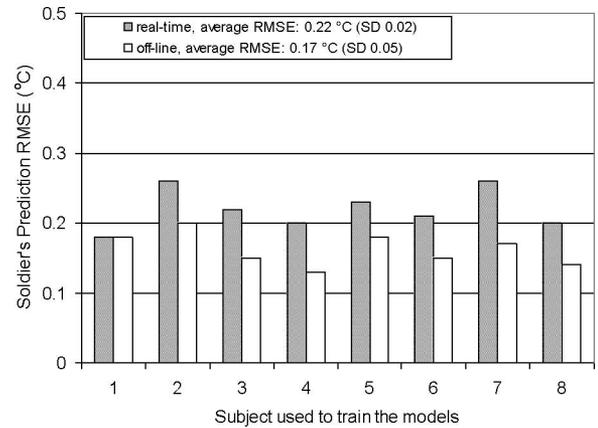


Fig. 3. Real-time and offline RMSEs for the cross-subject and cross-study 20-min-ahead predictions of the soldier's core temperature, using the models from each of the eight subjects of the first study.

In the third simulation, we compared the real-time and offline versions of the algorithms through cross-subject predictions, involving the two subjects of the second study. In this case, the cadet's entire core-temperature time-series data were predicted based on a model trained on the entire soldier's data and vice versa. We performed this test for two prediction horizons, 10 and 20 min, to investigate the dependency of the real-time algorithm on the prediction horizon. In this case, we expected the real-time algorithm to retain the properties of the offline version and yield a larger RMSE for the longer prediction horizon [4].

Figs. 4 and 5 show the results of the real-time algorithm for the cadet and the soldier, respectively, for both 10- and 20-min-ahead predictions. The two horizontal lines in the figures correspond to plausible physiological thresholds on human core temperature. The bottom line, at  $39\text{ }^{\circ}\text{C}$ , could be considered as an alert threshold warning that a person is exhibiting dangerously high levels of core temperature, while the top line, at  $39.5\text{ }^{\circ}\text{C}$ , could be taken as an alarm threshold indicating an imminent heat illness. This is supported by a clinical study that confirmed that 50% of the population will suffer at least a minor heat illness once the core temperature reaches  $39.5\text{ }^{\circ}\text{C}$  [11].

Analysis of the results in Fig. 1 shows that for the same-subject predictions, the average RMSE was 57% higher for the real-time algorithm than for its offline counterpart. However, the average RMSE of the real-time predictions at  $0.33\text{ }^{\circ}\text{C}$  was still rather small, suggesting that the performance degradation is acceptable and the technique is viable for real-time applications.

The results presented in Figs. 2 and 3 were very important from a model's universality and practical field application point of view. The results suggest that the average RMSEs obtained in the cross-subject and cross-study simulations were equivalent to those obtained for the same-subject predictions in Fig. 1. For example, the average RMSE of  $0.34\text{ }^{\circ}\text{C}$  for the cadet's real-time predictions in Fig. 2 was comparable to that of the same-subject predictions of  $0.33\text{ }^{\circ}\text{C}$  in Fig. 1. Interestingly, the average RMSE of  $0.22\text{ }^{\circ}\text{C}$  for the soldier's real-time predictions in Fig. 3 was considerably smaller than that of the same-subject predictions in Fig. 1. While these results indicate that the prediction errors

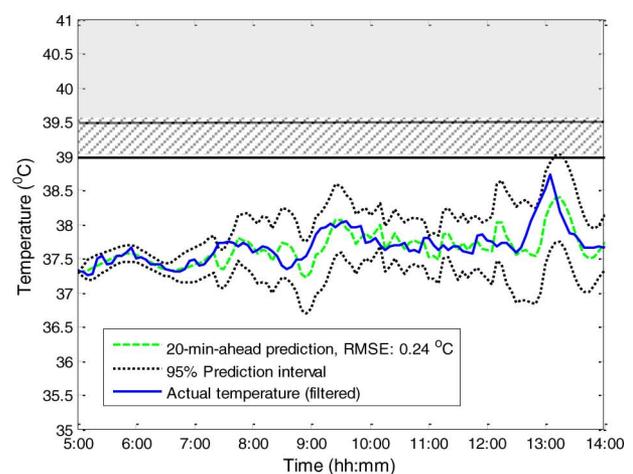
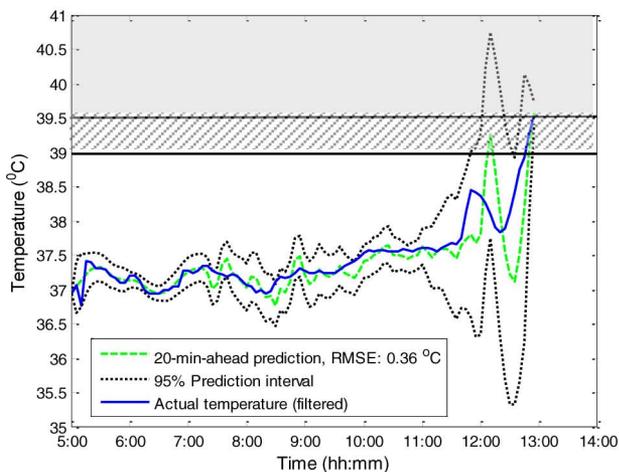
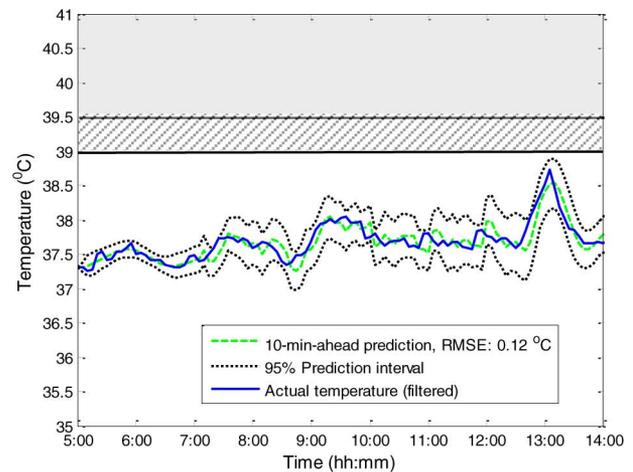
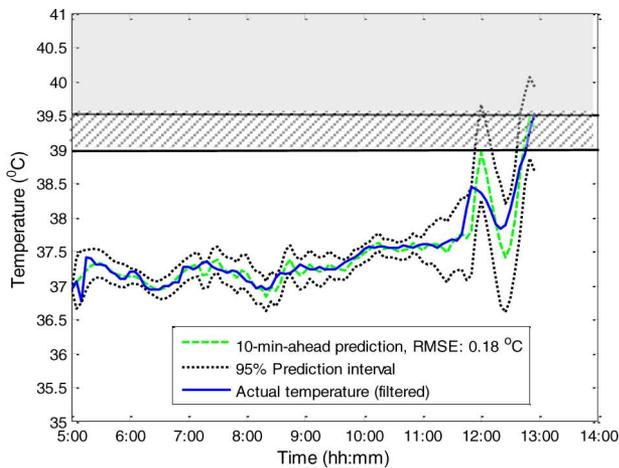


Fig. 4. Cadet predictions. 10-min-ahead predictions (top panel) and 20-min-ahead predictions (bottom panel), and their corresponding 95% PIs.

Fig. 5. Soldier predictions. 10-min-ahead predictions (top panel) and 20-min-ahead predictions (bottom panel), and their corresponding 95% PIs.

were somewhat dependent on the specific individuals used to train and test the models, they also indicate that, overall, the absolute values of the prediction errors were small. The real-time results, illustrated in Figs. 2 and 3, corroborated our previous finding using offline predictions [4] that AR models could be developed from one individual's data and subsequently used to predict other individuals.

The differences between real-time and offline RMSEs were caused by the so-called end effect observed in real-time filtering [8]. In real time, the filter can only use data up to time  $t$  to filter the signal at  $t$ . Conversely, offline, the whole time series is available, and future data are used to enhance the filtered data at time  $t$ , improving the performance of the AR model. This is particularly problematic in the prediction of oscillatory data, where the real-time filter, unlike its offline counterpart, cannot anticipate and correct for future, yet unknown curvatures in the data associated with upcoming inflection points. The end effect creates special and unique challenges for predictive algorithms, as the most recent samples, which carry the majority of the predictive information, cannot be properly filtered.

Notice that the prediction errors between the offline and real-time predictions were different for different subjects. This is

caused by the nature of the core-temperature signal being predicted. If the test core-temperature signal were smooth and exhibited very little variations, the difference between the offline and real-time versions would be very small. In the extreme, the two versions would produce identical results, if the test signal were a straight line. However, as the variability of the core-temperature signal increases, the two versions start to diverge, with the real-time version producing higher prediction errors due to the end effect of the real-time filtering. This is clearly noticed when we compare the offline and real-time prediction errors of the soldier (see Fig. 3) and the cadet (see Fig. 2). Because the soldier's data have less variability than the cadet's, his average RMSE increased by only 29% ( $0.17\text{ }^{\circ}\text{C}$  versus  $0.22\text{ }^{\circ}\text{C}$ ) from the offline to the real-time predictions, while that of the cadet increased by 54% ( $0.22\text{ }^{\circ}\text{C}$  versus  $0.34\text{ }^{\circ}\text{C}$ ). However, as pointed out earlier, the absolute errors were small.

Figs. 4 and 5 indicate that the point predictions for both subjects were quite accurate during the relatively stable portions of the core-temperature signal. However, due to the end effect issues, the point predictions, in particular, the ones with longer prediction horizons, did exhibit time lags in regions of large excursions of the core-temperature signal, for example,

~11:50 h for the cadet in Fig. 4 and ~13:10 h for the soldier in Fig. 5. The most interesting result, however, was that the algorithm was capable of predicting the dangerously high levels of the cadet's temperature at 12:55 h for both the 10- and 20-min prediction horizons. This was possible because the measured temperature follows a straight line from 12:20 h onward. Actually, as illustrated in Fig. 4, the cadet's predictions reached the alert threshold of 39°C much earlier, at ~12:10 h, providing an early indication of the dangerous trend in core temperature. The significant overprediction of the measured temperature at ~12:10 h is perhaps less important than the observation that the algorithm was capable of forecasting a tendency of sharp increases in the cadet's core temperature. In contrast, it is worth noticing that the soldier's predictions never crossed the alert threshold, indicating that the algorithm was able to correctly predict the more stable nature of his core temperature.

Analyses of the 95% PIs in Figs. 4 and 5 show that they reflected the expected uncertainty in core-temperature predictions, as they became significantly wider in the regions with larger signal variations. Also, as expected, the PIs were wider for longer prediction horizons (bottom panels) because the confidence in the predictions decreases with increasing horizons. We should also point out that, when compared with the offline computations [5], the real-time PIs were generally wider, reflecting the larger uncertainty of the real-time predictions. As illustrated in Fig. 4, the PIs could also be used as warning mechanisms, since the cadet's PIs crossed the alert and alarm thresholds even earlier than the point predictions. Another important observation, as shown in the bottom panel in Fig. 4, is that the predictions at ~12:55 h made 20 min earlier possessed very small uncertainties, i.e., had narrow PIs, indicating that the algorithm was rather confident that the cadet's core temperature was reaching the 39.5°C alarm threshold. The larger uncertainty for the same few points obtained with the 10-min-ahead predictions (see top panel in Fig. 4) is attributed to the larger noise level of the samples around 12:45 h (not apparent in Fig. 4). However, as mentioned earlier, on average, the 10-min-ahead predictions were more accurate than the 20-min-ahead predictions. It is also reassuring that the PIs for the soldier in Fig. 5 remained under the alert threshold practically all the time, barely crossing it at ~13:00 h for the 20-min-ahead predictions. This fact also indicates that the soldier's core temperature remained stable and regulated during the whole exercise.

The comparison of the RMSEs for 10- and 20-min prediction horizons in Figs. 4 and 5 revealed that doubling the length of the prediction horizon effectively doubled the RMSE. This observation suggests that the RMSE may be a linear function of the prediction horizon. The cadet's RMSE was higher than the soldier's RMSE for both the prediction horizons, indicating that prediction of a more volatile core-temperature signal was less accurate than the prediction of a more stable one. This is consistent with the results illustrated in Figs. 2 and 3. These and our earlier results [4] also indicate that the most accurate predictive models are those developed using the most encompassing training data, involving the widest-possible variations of the core-temperature data. Hence, the ideal training data should consist of long records, preferably longer than 24 h, containing

the full range of expected core-temperature variations, including extreme and dangerous values. Accordingly, in our case, the universal model should be based on the cadet's core-temperature data.

A very encouraging aspect of this study is that the universality of the predictive models has been preserved, and the conclusions reached previously for the offline version of the algorithm [4], stating the possibility of cross-subject and cross-study predictions, and, hence, model universality, also holds for the real-time version. Our conclusions concerning the better generalization capabilities of regularized models were also confirmed, as in our real-time simulations, the regularized models produced more accurate and stable predictions (results not shown).

The end effect in the real-time filtering of the raw signal is undoubtedly the largest limitation of the proposed algorithm, as it introduces lags in the predictions of oscillatory data, effectively decreasing the prediction horizon. Another limitation is the use of the computationally expensive bootstrap method for real-time estimation of PIs in resource-limited wearable devices.

Simulation results with field data suggest that the real-time implementation of the core-temperature prediction algorithm could be a valuable tool for early warning of an impending heat illness in humans. Although not as accurate as the offline algorithm, the real-time implementation yielded forecasts with sufficient fidelity for practically meaningful prediction horizons. The results also suggest that the decrement in prediction accuracy could be compensated by the incorporation of alert and alarm thresholds in core temperature. For the cadet, the alarm threshold was reached well before the measured core temperature underwent a continuous sharp increase, whereas for the soldier, the lower (alert) threshold was barely reached.

Importantly, the universality of the data-driven models has been preserved, indicating that models could be developed offline from one individual's data and applied to predict the core temperature of other individuals in real time. Moreover, the previously developed PIs, based on the bootstrap method, placed around the predictions provided a useful and intuitively correct measure of the reliability of the core-temperature point predictions.

#### IV. CONCLUSION

The real-time core-temperature prediction algorithm is currently being implemented as part of a physiologic monitoring system for dismounted military personnel, where the cadet's data are being used to construct a universal model offline for real-time predictions of other individuals. The algorithm will undergo extensive field tests and, if proven successful, could become an important warning tool for impending heat illnesses during strenuous physical activities in hot-weather conditions. Our ongoing research efforts are focused on improving the real-time filtering of the raw signal to better address the end-effect problem and implementing an analytical and less computationally expensive expression for estimating statistically based PIs in real time.

## DISCLAIMER

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution.

## REFERENCES

- [1] D. S. Moran, Y. Heled, L. Still, A. Laor, and Y. Shapiro, "Assessment of heat tolerance for post exertional heat stroke individuals," *Med. Sci. Monit.*, vol. 10, pp. CR252–CR257, Jun. 2004.
- [2] D. S. Moran, A. Shitzer, and K. B. Pandolf, "A physiological strain index to evaluate heat stress," *Amer. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 275, pp. 129–134, 1998.
- [3] V. S. Miller and G. P. Bates, "The thermal work limit is a simple reliable heat index for the protection of workers in thermally stressful environments," *Ann. Hyg.*, vol. 51, no. 6, pp. 553–561, 2007.
- [4] A. V. Gribok, M. J. Buller, and J. Reifman, "Individualized short-term core temperature prediction in humans using biomathematical models," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 5, pp. 1477–1487, May 2008.
- [5] A. V. Gribok, M. J. Buller, R. W. Hoyt, and J. Reifman, "Providing statistical measures of reliability for body core temperature predictions," in *Proc. Conf. IEEE Eng. Med. Biol. Soc.*, 2007, vol. 1, pp. 545–548.
- [6] F. Gustafsson, "Determining the initial states in forward-backward filtering," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 988–992, Apr. 1996.
- [7] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. New York: Wiley, 1977.
- [8] C. Chatfield, *Time-Series Forecasting*. Boca Raton, FL: Chapman & Hall/CRC Press, 2001.
- [9] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. London, U.K.: Chapman & Hall, 1993.
- [10] C. Morris, G. Atkinson, B. Drust, K. Marrin, and W. Gregson, "Human core temperature responses during exercise and subsequent recovery: An important interaction between diurnal variation and measurement site," *Chronobiol. Int.*, vol. 26, pp. 560–575, 2009.
- [11] R. F. Goldman, "Introduction to heat-related problems in military operations," in *Medical Aspects of Harsh Environments*, K. B. Pandolf and R. E. Burr, Eds. Falls Church, VA: Office of the Surgeon General, U.S. Army, 2001, pp. 3–49.



**Andrei V. Gribok** received the B.S. degree in systems science and the M.S. degree in nuclear engineering from Moscow Institute of Physics and Engineering, Moscow, Russia, both in 1987, and the Ph.D. degree in biological physics from Moscow Institute of Biological Physics, Moscow, in 1996.

He was engaged with the Telemedicine and Advanced Technology Research Center, Fort Detrick, MD, on an interpersonal agreement assignment. He is currently a Research Assistant Professor in the Department of Nuclear Engineering, The University of Tennessee, Knoxville. His research interests include inverse and ill-posed problems in engineering, statistical learning, and model misspecification in statistics.



**Mark J. Buller** received the B.Sc. (Hons.) degree in applied psychology from the University of Wales College of Cardiff, Cardiff, U.K., in 1991, and the M.Sc. degree in computer science from Brown University, Providence, RI, in 2008. He is currently working toward the Ph.D. degree in computer science at Brown University.

He is also with the U.S. Army Research Institute of Environmental Medicine, Natick, MA. He has been involved as a Lead in the development of wearable physiological monitoring systems for over ten years, and is a member of the North Atlantic Treaty Organization Research Technology Group "Real-Time Physiological and Psycho-Physiological Status Monitoring for Human Protection and Operational Health Applications." His research interests include understanding human health state in harsh environments through the use of machine learning techniques.



**Reed W. Hoyt** received the B.S. degree in biology from the University of Arizona, Tucson, in 1972, and the Ph.D. degree in biomedical sciences with a specialization in physiology from the University of New Mexico, Albuquerque, in 1981.

He is the Division Chief of the Biophysics and Biomedical Modeling Division, U.S. Army Research Institute of Environmental Medicine, Natick, MA. His current research interests include the development and use of ambulatory physiological monitoring technologies to study soldiers and first responders, and the development of new methods to understand the effects of exercise and the environment on the energy, water, and metabolic fuel requirements of humans.



**Jaques Reifman** received the B.S. degree in civil engineering from Rio de Janeiro State University, Rio de Janeiro, Brazil, in 1980, the B.B.A. degree in business administration from Rio de Janeiro Federal University, Rio de Janeiro, in 1985, and the M.S.E. and Ph.D. degrees in nuclear engineering from the University of Michigan, Ann Arbor, in 1985 and 1989, respectively.

He is currently a Senior Research Scientist in the Department of the Army, U.S. Army Medical Research and Materiel Command (USAMRMC), Fort Detrick, MD, where he is also the Founder and the Director of two organizations: the Department of Defense Biotechnology High Performance Computing Software Applications Institute for Force Health Protection and the USAMRMC Bioinformatics Cell at the Telemedicine and Advanced Technology Research Center. His current research interests include the areas of physiological signal processing, statistical pattern recognition, artificial intelligence, data mining, biomathematical modeling, systems biology, bioinformatics, genomics, and proteomics.