Project Report
PSM-4

Validation of Core Temperature Estimation Algorithm

K.M. Showers
A.R. Hess
B.A. Telfer

TBD 2016

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Distribution A: Public Release
Prepared for the U.S. Army Medical Materiel Development Activity (USAMMDA)
under Air Force Contract FA8721-05-C-0002.

This report is the result of studies performed at Lincoln Laboratory, a federally funded research and development center operated by Massachusetts Institute of Technology. This material is based on work supported by the U.S. Army Medical Materiel Development Activity (USAMMDA) under Air Force Contract No. FA8721-05-C-0002. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Department of the Air Force.

Massachusetts Institute of Technology
Lincoln Laboratory

Validation of Core Temperature Estimation Algorithm

K.M. Showers
A.R. Hess
B.A. Telfer

Project Report PSM-4

TBD 2016

Lexington Massachusetts

This page intentionally left blank.

DISCLAIMERS

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

Citations of commercial organizations and trade names in this report do not constitute an official U.S. Army endorsement or approval of the products or services of these organizations.

This page intentionally left blank.

ABSTRACT

Real-time knowledge of core body temperature is a key input to determine thermal-work strain and risk of heat injury. An algorithm for estimating core temperature based on heart rate has been developed by others in order to avoid standard but more invasive measurements, such as ingestible capsules and rectal or esophageal probes. This report provides an independent assessment of the algorithm, based on both parametric variations and field data that were not used in algorithm development. Assessment through parametric variation shows qualitatively expected behavior. Field data were taken from a study of 33 young male military personnel who tested six prototype uniforms over the course of 12 days in Okinawa, Japan. The field data selected for assessment consist of nearly 48,000 measurements of heart rate and core temperature. Core temperature was measured by an ingestible capsule. The observed core temperature range was 36.1–39.5°C. Bland–Altman analysis yielded a bias of –0.01°C and a 95% limit of agreement (LoA) of ±0.58°C. The LoA is similar to the ±0.5°C LoA resulting from variations in ingestible capsule temperatures, and is also consistent with the ±0.63°C LoA found by the algorithm developer on other field data. The accuracy of the Physiological Strain Index (PSI), a standard estimate of thermal-work strain on a nominal 0–10 scale, was also assessed with the estimated core temperature as an input. A bias of –0.01 and LoA of ±1.2 was found. The PSI LoA is comparable to the LoA resulting from the variability in computing PSI with two different ingestible capsule temperatures. Overall, within the applicability of the field data that were evaluated, the core temperature estimation algorithm is sufficiently accurate and precise to provide a field-expedient indication of core temperature and thermal-work strain. Follow-on independent assessments of the algorithm are recommended as additional field data become available for higher core body temperatures and from women. This report focuses on evaluating the algorithm for accuracy in estimating core temperature and PSI. Quantifying the sensitivity and specificity of these measures for predicting heat injury remains for future work.

This page intentionally left blank.

ACKNOWLEDGEMENTS

The sponsorship of this work by the U.S. Army Medical Materiel Development Activity is gratefully appreciated. We thank Mr. Mark Richter from the Marine Expeditionary Rifle Squad for providing the large data set for evaluation. We appreciate the careful review of this report by Dr. Paula Collins, Dr. Jeffrey Palmer, and Dr. James Williamson.

This page intentionally left blank.

TABLE OF CONTENTS

		Page
	Disclaimers	iii
	Abstract	v
	Acknowledgements	vii
	List of Illustrations	xi
1.	INTRODUCTION	1
2.	ALGORITHM BACKGROUND	3
	2.1 Expectations for Core Temperature Accuracy	3
	2.2 Core Temperature Estimation Algorithm	3
3.	METHODS	5
	3.1 Overview	5
	3.2 Parametric Analysis	5
	3.3 Field Data Analysis	5
4.	RESULTS	9
	4.1 Parametric Variation	9
	4.2 Field Data	9
5.	DISCUSSION	13
6.	LIMITATIONS	15
7.	CONCLUSION	17
8.	REFERENCES	19

TABLE OF CONTENTS
(Continued)

Page

Appendix A. Weather Data 21

Appendix B. Breakdown of Selected Data 23

Appendix C. Breakdown of RMSE Values 25

Appendix D. Core Temperature Estimation Algorithm 27

LIST OF ILLUSTRATIONS

Figure No.		Page
1	Normalized histogram of observed core temperatures.	7
2	Core temperature estimation from parametrically varying heart rate: (a) time required for the algorithm to converge to within 0.5% of the final core temperature, (b) converged temperature.	9
3	(a) Scatter plot of observed versus estimated core temperature with the line of identity (dashed) and the least squares regression line (solid) and line equation in the top left corner. (b) Bland–Altman plot showing the bias (center line) and LoA (top and bottom lines). (c) Normalized histogram of error between observed and estimated temperatures.	10
4	Bland–Altman plots with x-axis as (a) gold standard and (b) mean of observed and estimated core temperatures.	10
5	Normalized histogram of core temperature estimation (a) bias and (b) RMSE calculated for each subject and day.	11
6	Example of a subject and day with a (a) low RMSE (0.10 for subject 23, day 5) and b) high RMSE (0.67 for subject 31, day 7).	11
7	(a) Scatter plot of observed versus estimated PSI with the line of identity (dashed) and the least squares regression line (solid) and line equation in the top left corner. (b) Bland–Altman plot showing the bias and LoA showing the bias (center line) and LoA (top and bottom lines). (c) Normalized histogram of error between observed and estimated PSI.	12
A-1.	(a) High and low temperatures for each study day. (b) Average humidity.	21
B-1.	Duration of data for each subject and day that were used in the analysis. Green indicates greater than 2 hours of data collected; yellow, 0–2 hours; red, 0 data points. Blue highlights subject and day high and low.	23
C-1.	RMSE value for each subject and day. Green coloring indicates lowest 10% of RMSE values. Red indicates highest 10% of RMSE values.	25

This page intentionally left blank.

1. INTRODUCTION

Real-time knowledge of core body temperature is a key input to determine thermal-work strain and risk of heat injury. Accepted standards for measuring core temperature include probes in the pulmonary artery, rectum, or esophagus, and an ingestible capsule that transmits the temperature to an external receiver as the capsule travels through the digestive tract. Of these, only the ingestible capsule is suitable for use for outside the clinic or laboratory, but the number of capsules required and their cost make this option impractical for routine use. Thus, there is a need to estimate core temperature from noninvasive sensor measurements that can be made in real time in the field. The U.S. Army Research Institute of Environmental Medicine (USARIEM) has developed such an estimation algorithm [1], which is based on heart rate. The algorithm's intent is not to replace clinical measurements but to provide a field-expedient solution for estimating core temperatures.

This report provides an independent assessment of the accuracy of the estimation algorithm on a large set of field data that was not used by USARIEM in developing the algorithm and that has not been previously assessed.

This page intentionally left blank.

2. ALGORITHM BACKGROUND

2.1 EXPECTATIONS FOR CORE TEMPERATURE ACCURACY

To understand the level of accuracy needed from a core temperature measurement or estimate, it is helpful to consider the ranges of measurements involved, as well as the variance in gold standard measurements.

For thermal strain, the range of core body temperatures of interest is about 3°C. A normal resting temperature is about 37°C and can rise up to about 40°C, which is a common symptom of exertional heat stroke [2], although not a critical temperature [3]. Given this small range, an estimate of core body temperature should have an accuracy of well under 1°C to be meaningful. A lower bound for the expected accuracy is about ±0.5°C, which encompasses 95% of the variations observed between ingestible capsule temperatures as the capsules descend through the digestive tract [4].

To assess thermal-work strain, core temperature is incorporated into a physiological strain index (PSI) [5], given by

$$PSI = 5 \cdot \frac{HR - HR_0}{180 - HR_0} + 5 \cdot \frac{CT - CT_0}{39.5 - CT_0} \tag{1}$$

where HR is the heart rate, CT is the core temperature, and HR_0 and CT_0 are the initial resting values. An error in CT of 0.5°C will result in a PSI error of about 1, depending on the specific value of CT_0.

2.2 CORE TEMPERATURE ESTIMATION ALGORITHM

The core temperature estimation algorithm is based solely on heart rate. The physiological basis for this is twofold. First, heart rate increases with work, which heats the body core. Muscles are only about 20% efficient, with 80% of the energy generated during work going to heat production [6]. Second, heart rate increases to support the body's heat dissipation. To dissipate heat, blood vessels near the skin vasodilate to increase blood perfusion. Thus, heart rate increases both to support the cardiac output needed both to perform work and to increase skin blood flow to allow dissipation of the resulting heat.

To model these relationships, the core temperature estimation algorithm empirically defines a dynamic relationship between heart rate and core temperature. This algorithm is based on an extended Kalman filter, which was developed using field data from 17 young male U.S. Army soldiers with core temperatures ranging from 36–40ºC [1]. The algorithm is initialized with CT_0 and updated with heart rate measured at one minute intervals. The initial core temperature can be measured at the start of an exercise or assumed to be 37.1ºC. Properly treating the core temperature initialization is important for assessing algorithm accuracy. This topic is addressed later in this and in Section 3.3.2. An additional parameter for initial model variance represents the confidence in the initial core temperature measurement, which has been selected as 0.01 [1]. MATLAB code used to evaluate the core temperature estimation algorithm is listed in Appendix D.

To understand the applicability of the algorithm, it is helpful to understand the range of data that the developer used to evaluate the algorithm. USARIEM has evaluated the algorithm accuracy on over 52,000 core temperature measurements from nine laboratory and field studies with a total of 82 male and one female volunteers with mean age ranging from 22 to 28 years [1]. A wide range of work rates (estimated as 200–1000 W), air temperatures (9–47°C), and relative humidities (9–95%) were represented. The data sets also included various hydration states, clothing ensembles, and acclimatization states. Core temperature was measured by rectal probe, thermometer pill suppository, and ingested thermometer pill. For the entire set of validation data, a bias of –0.03°C and 95% limit of agreement of ±0.63°C were measured.

The algorithm has also been evaluated by the developers on data from 25 male and two female soldiers (age = 30 ± 6 years) from three training events for responding to chemical and biological hazards [7]. Encapsulating personal protective equipment was worn. Fewer than 2,000 data points were available. A bias of –0.02°C and limit of agreement of ±0.48°C were measured.

In these algorithm assessments by the developer, the initial core temperature was set to the observed temperature. This was found to result in optimistic estimation accuracies for about the first 30 minutes of data, as opposed to setting the initial temperature to an a priori value, e.g., 37.1°C [7]. Thus for the purposes of this analysis, it is important in assessing accuracy to not include the first 30 minutes of temperature estimates after initialization, so that the temperature estimates are not unfairly influenced by the initial temperature.

3. METHODS

3.1 OVERVIEW

The algorithm assessment is based on both parametric analysis and field data. The parametric analysis allows the algorithm performance to be examined across the entire space of possible inputs to detect any undesired behavior, such as discontinuities. The parametric analysis is also valuable for providing insight into the algorithm behavior. Assessment of field data is critical for determining performance under real world conditions compared to a gold standard.

3.2 PARAMETRIC ANALYSIS

For parametric analysis, an initial resting condition is set as $CT_0 = 37.1^{\circ}$C and initial confidence of 0.01. The algorithm is then run on a constant heart rate until the *CT* estimate converges, with convergence defined as the time for *CT* to fall within 0.5% of the final temperature. Time to convergence and temperature at convergence are plotted.

3.3 FIELD DATA ANALYSIS

3.3.1 Field Test Description

The Marine Expeditionary Rifle Squad provided de-identified data from a study conducted at the U.S. Marine Corps Jungle Warfare Training Center, Okinawa, Japan from June 17–19 and 21–29, 2013, for a total of 12 days. The study was approved by the cognizant institutional review board. Thirty-three male military personnel aged 21.8 ± 3.3 (mean ± one standard deviation) years participated. Body weight and height were 81.3 ± 9.6 kg and 174.4 ± 5.7 cm, respectively. The study's purpose was to test six prototype uniforms for suitability and comfort in a subtropical jungle. The subjects were rotated through the uniform prototypes so that each subject wore each prototype at least once. The study was conducted during both day and night for varying amounts of time and activity.

During the study, heart rate and core temperature were recorded at 15 second intervals using a physiological status monitor (PSM, Equivital EQ-02 LifeMonitor, Hidalgo, Cambridge UK) and ingested thermometer capsule (Respironics, Bend, OR).

Based on nearby Kadena Air Base measurements, the mean temperature was 28.9°C throughout the study, with a high of 30.6 ± 0.43°C and a low of 26.7 ± 1.0°C. On June 18th, 19th, and 25th, (days 2, 3, and 8) precipitation was 2.7, 0.3, and 0.5 centimeters, respectively. Additional weather information can be found in Appendix A. Weather Data.

3.3.2 Data Selection

To identify valid data for algorithm assessment, the data were preprocessed in several steps.

First, core temperature data were associated with particular subjects by analyzing the pill serial numbers that were received by each subjects's PSM. This process eliminated any core temperature crosstalk from other subjects, which can occur when subjects are within a few meters of each other.

Second, core temperature data were manually reviewed for indications that eating or drinking disturbed the measurements. This was indicated by sharp declines in core temperature readings followed by a slow return (up to an hour) to a baseline [7]. These and other periods of noncredible temperature data were manually removed from the analysis.

Third, heart rate was recomputed based on the interbeat intervals recorded by the PSM, since the recorded heart rates are susceptible to motion artifacts. The recomputation involved median filtering and additional outlier rejection. As a final step, the recomputed heart rate measurements were manually reviewed for accuracy. Noncredible, erratic, and nonphysiological measurements were removed from the analysis.

After removing poor-quality data, the total amount of data available for each subject and day varied considerably. Totaled by subject, data duration ranged from 1.4 to 40 hours. Totaled by day, data duration ranged from 30 to 118 hours. Combining all subjects and all days, the total data consisted of 47,549 observations (792.5 hours, with one data sample for each minute). Although this corresponds to only 46% of the total core temperature data collected, the number of data points is quite large, essentially equal to the number of data points from the nine studies that were evaluated in [1]. A distribution of data duration per subject and per day is provided in Appendix B. Breakdown of Selected Data.

A histogram of the available observed core temperatures readings can be seen in Figure 1. Of the 47,549 data points that were used in the validation, 136 (0.3%) data points of the observed core temperature readings were above 39°C. Of the 136 observed temperature readings above 39°C, 61 occurred in Subject 31 on Day 6 and 7.

3.3.3 Core Temperature Estimation

The core temperature estimation algorithm operates on heart rate values provided at one-minute intervals. To match this time interval, one-minute medians were computed of the 15 s heart rate and observed core temperatures. The estimation algorithm must also be initialized with a core temperature and model variance, which were set to 37.1°C and 0.01, as was done in [7].

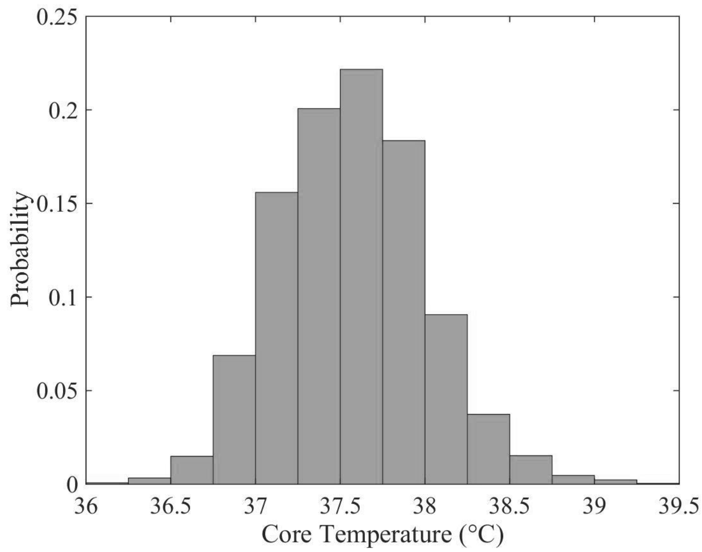


Figure 1. Normalized histogram of observed core temperatures.

During the data selection, gaps of varying durations were introduced into the heart rate and core temperature data. For time gaps less than 10 minutes, model parameters (estimated temperature and model variance) were carried over from the previous time resulting in a continuation of the algorithm as if no gap existed. For time gaps of greater than 10 minutes, the algorithm was reinitialized with a temperature of 37.1ºC and model variance of 0.01. The true core temperature as measured by the ingestible capsule can differ from this initial value by 1°C or more, depending on the individual, time of day, and activity level. So that the algorithm was not penalized for start-up errors, the first 30 minutes of core temperature estimates were discarded each time the algorithm was reinitiated. The value of 30 minutes was chosen based on an analysis of algorithm convergence time.

3.3.4 Statistics

Accuracy of the core temperature estimation algorithm and PSI were assessed using the Bland–Altman method [8], which computes a bias and limit of agreement (LoA) as the mean and ±1.96 standard deviations of the estimated minus observed core temperatures. The LoA provides a range of error within which 95% of the estimated errors should fall assuming a normal distribution, which is consistent with the error distribution reported in Section 4.

The Bland–Altman plot shows the relationship between the estimated and observed values by plotting the difference on the y-axis. On the x-axis, either the average of the two values or the gold standard value, as recommended by [9], can be plotted. Both plots were generated for comparison.

The root mean squared error (RMSE) was also computed, as given by Equation 2.

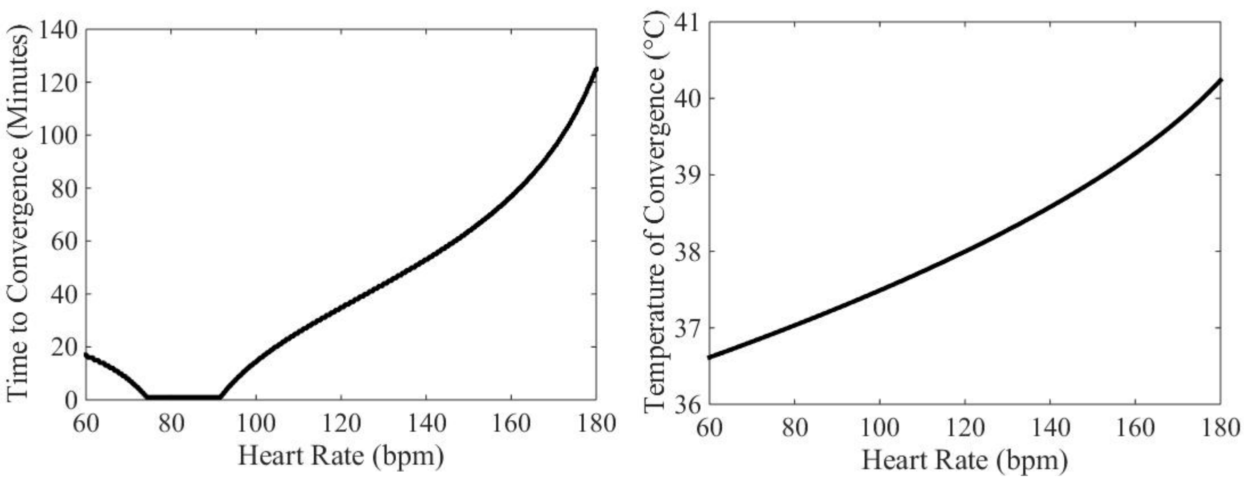
$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(Temp\ Obs - Temp\ Est)^2}{N}} \tag{2}$$

4. RESULTS

4.1 PARAMETRIC VARIATION

Consider a case in which a person starts exercising from rest (let CT_0 = 37.1°C) and maintains a steady heart rate. The core temperature estimate will converge to a value plotted in Figure 2(a), with the time to convergence given in Figure 2(b). Convergence time is defined as the time for temperature to converge within 0.5% of the true steady-state value.

An example result is that if a person maintains a heart rate of 160 beats per minute (bpm), the core temperature estimate will converge to 39.3°C after 77 minutes. At heart rates of 74 to 92 bpm, the estimate converges within one update to a temperature close to the initial temperature of 37.1°C. For heart rates below 83 bpm, the estimate converges to temperatures below the initial temperature, but still within the normal range. All of these behaviors are expected based on algorithm inspection, and no unexpected discontinuities are observed.



(a) (b)

Figure 2. Core temperature estimation from parametrically varying heart rate: (a) time required for the algorithm to converge to within 0.5% of the final core temperature, (b) converged temperature.

4.2 FIELD DATA

Figure 3 depicts the algorithm accuracy through a scatter plot, Bland–Altman plot, and histogram of errors. In the Bland–Altman plot of Figure 3(b), the observed core temperature is used as the abscissa. The bias is –0.01°C with LoA of ±0.58°C. Figure 3(c) is a normalized histogram of the error between observed and estimated temperatures.

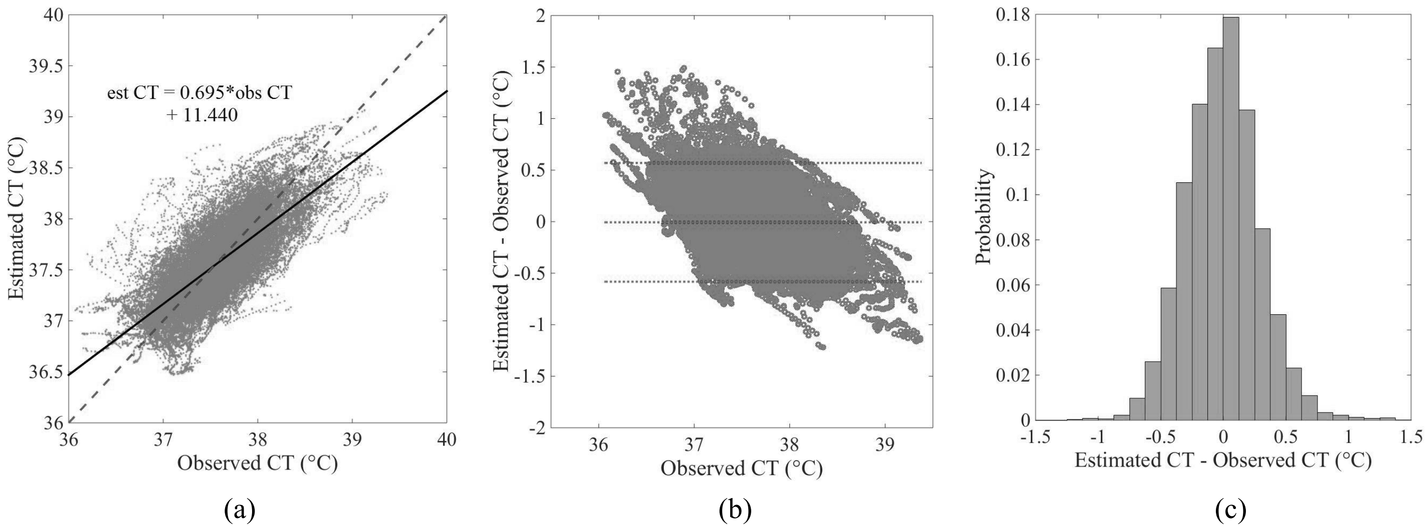


Figure 3. (a) Scatter plot of observed versus estimated core temperature with the line of identity (dashed) and the least squares regression line (solid) and line equation in the top left corner. (b) Bland–Altman plot showing the bias (center line) and LoA (top and bottom lines). (c) Normalized histogram of error between observed and estimated temperatures.

Since interpretation of the Bland–Altman plot may differ depending on choice of the abscissa, Figure 4 compares the two plot variations (Krouwer, 2008). Figure 4(a) shows the gold standard on the x-axis while Figure 4(b) shows the mean of the estimated and observed core temperature.

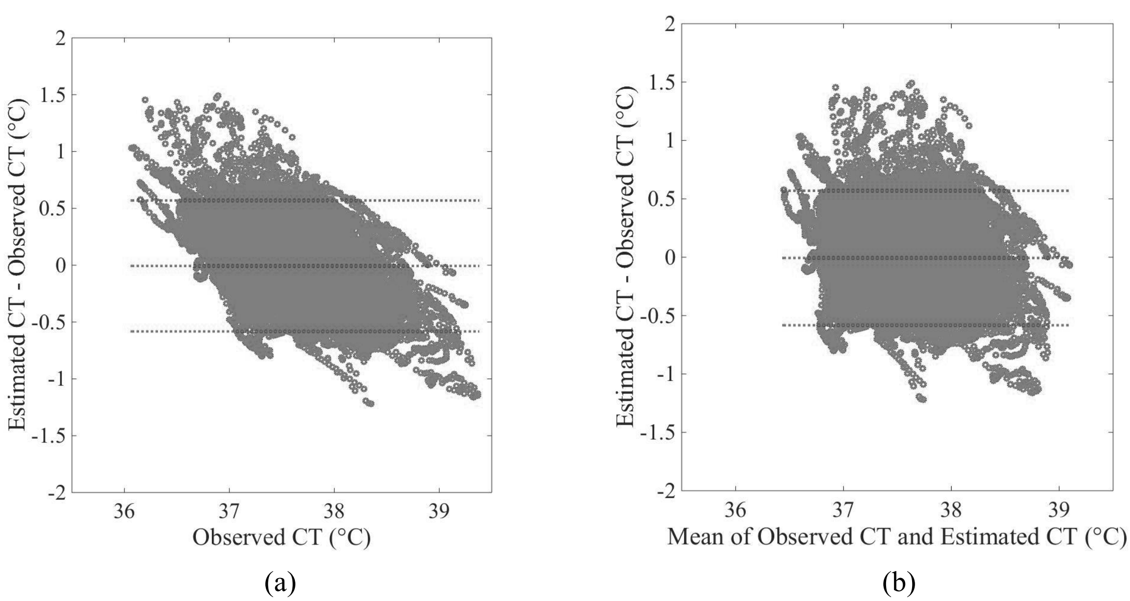
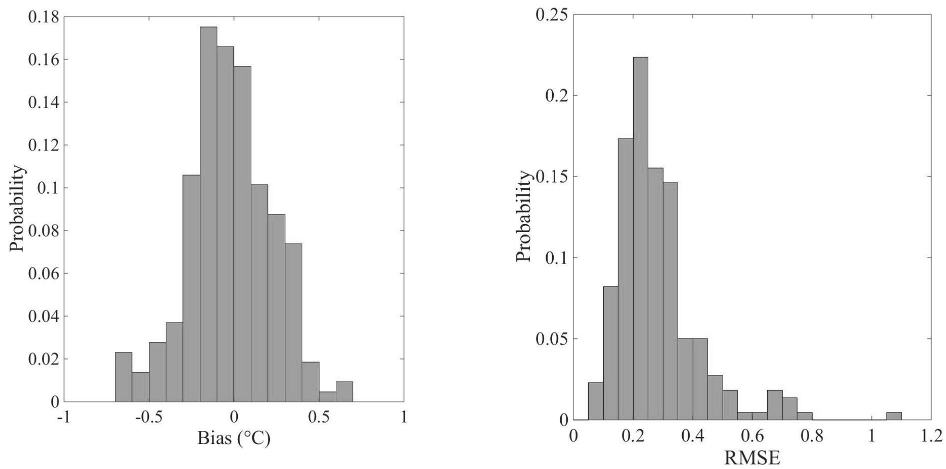


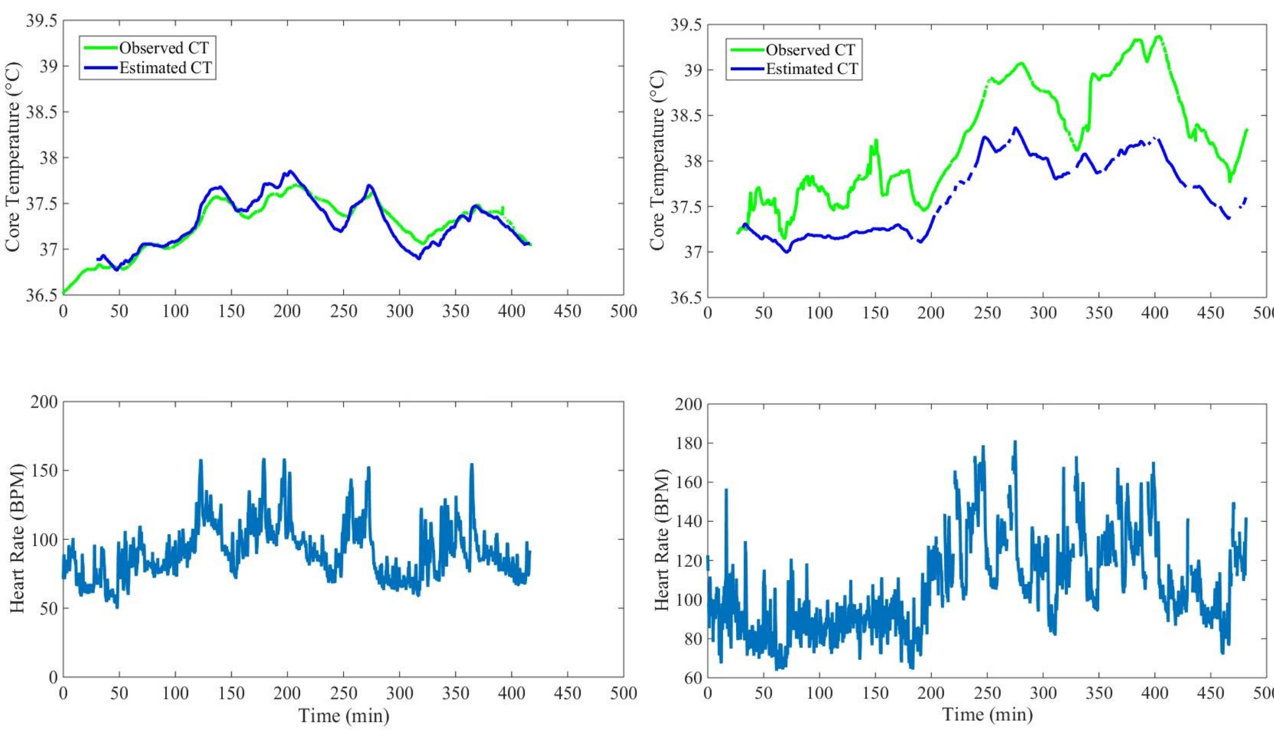
Figure 4. Bland–Altman plots with x-axis as (a) gold standard and (b) mean of observed and estimated core temperatures.

A bias and an RMSE value were calculated for each subject for each day of the collection. A histogram of the resulting values can be seen in Figure 5. The overall bias was –0.01 and RMSE was 0.29. A further breakdown in RMSE values for each subject on each day can be found in Appendix C.



(a) (b)

Figure 5. Normalized histogram of core temperature estimation (a) bias and (b) RMSE calculated for each subject and day.



(a) (b)

Figure 6. Example of a subject and day with a (a) low RMSE (0.10 for subject 23, day 5) and b) high RMSE (0.67 for subject 31, day 7).

Figure 6 gives examples of subjects and days with high and low RMSE values. Figure 6(a) has an RMSE and bias of 0.10°C and –0.01°C, respectively. Figure 6(b) has an RMSE and bias of 0.22°C and –0.61°C, respectively.

Short time gaps (<10 minutes) are visible in the estimated *CT* in Figure 6(b). These gaps are due to inaccurate heart rate data that were excluded from the analysis. The core temperature estimates were affected slightly when these short gaps occured. The gaps were handled by carrying over model parameters from the beginning of the gap, resulting in a continuation of the algorithm as if no gap existed. If the gap had not existed, then additional heart rate data would have been available that would have updated the temperature estimate. If the true core temperature were rising during the gap, then the gap would result in a slightly lower estimated temperature than if the gap had not existed. To gauge the effect of these gaps on accuracy, the heart rates in the gaps in the Figure 6(b) example were linearly interpolated and the estimated core temperatures were recomputed. This alternate approach resulted in estimated temperatures that were slightly higher, by up to about 0.1°C. Because this was not a large effect, the algorithm was assessed only using measured, and not interpolated, heart rates.

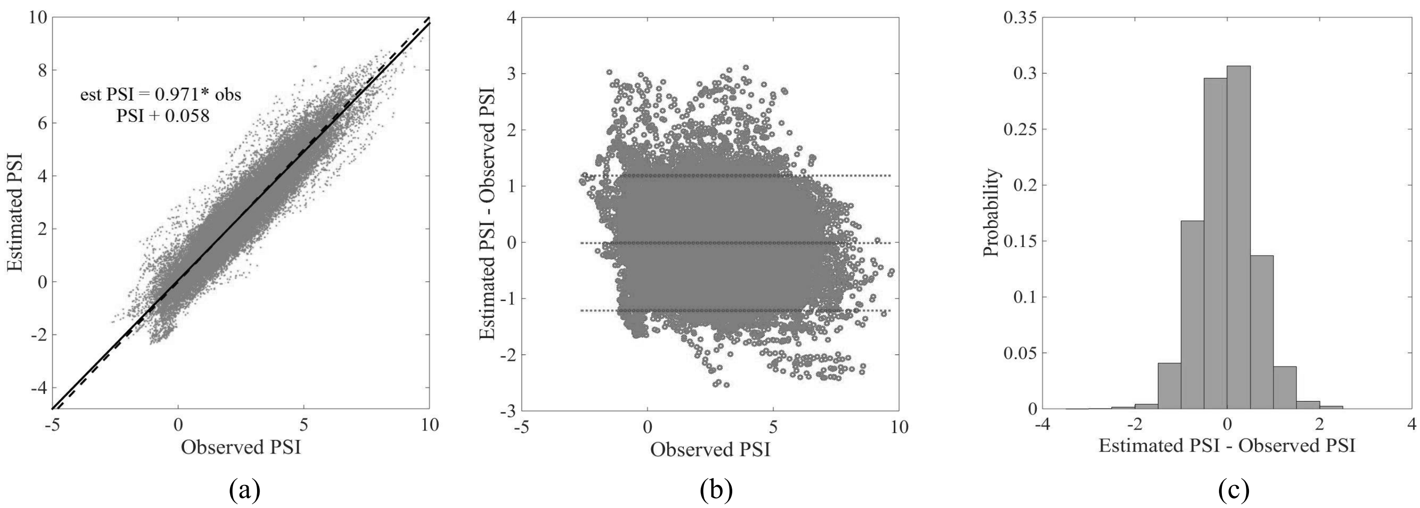


Figure 7. (a) Scatter plot of observed versus estimated PSI with the line of identity (dashed) and the least squares regression line (solid) and line equation in the top left corner, (b) Bland–Altman plot showing the bias and LoA showing the bias (center line) and LoA (top and bottom lines), (c) normalized histogram of error between observed and estimated PSI.

The PSI was calculated using observed and estimated core temperatures, with CT_0 = 37.1°C and HR_0 = 71 bpm [5]. The PSI has a bias and LoA of –0.01 ± 1.20, with an RMSE of 0.61. These values are illustrated in Figure 7. The negative PSI values in Figure 7 are due to data where $CT < CT_0$ or $HR < HR_0$.

5. DISCUSSION

The accuracy of a core temperature estimation algorithm has been independently assessed, based on both parametric variations and on field data that were not used in algorithm development. Parametric variation revealed no unexpected behavior. Assessment of field data yielded a negligible bias of -0.01°C and a LoA of $\pm 0.58^{\circ}$C. The LoA is similar to variations seen in gold standard measurements, such as the LoA of $\pm 0.5^{\circ}$C for variations in ingestible capsule temperatures [4]. The LoA is also consistent with LoAs found by the algorithm developer from evaluation of other field data, namely $\pm 0.63^{\circ}$C [1] and $\pm 0.48^{\circ}$C [7].

Whether the Bland–Altman plots exhibit any correlation of error with temperature depends on the choice of abscissa. With the abscissa as the gold standard measurement (Figure 3b), the estimated temperature tends to be too high for observed temperatures below 37°C and too low for observed temperatures above about 38.5°C. This correlation is also visible in the Figure 3(a) scatter plot in the misalignment of the line of identity and the least squares regression line. When the x-axis of the Bland–Altman plot is instead chosen as the mean of the observed and estimated data, as in Figure 4(b), the correlation is no longer apparent. This is expected, since at low observed temperatures, the higher estimated temperatures will increase the mean values, and at high observed temperatures, the lower estimated temperatures will decrease the mean values. Both of these effects will tend to make the correlation less visible. The correlation of error with temperature seen in Figure 3(b) must be interpreted somewhat tentatively, given the relatively limited number of data points with observed core temperatures above 38.5°C, and particularly above 39°C.

From examining core temperature estimation RMSE values broken down by subject and day in Appendix C, no strong consistency in errors is seen across subjects. A subject's RMSE may be relatively high on some days and relatively low on other days, but they are not consistently high or low. Across days, day 5 appears to have lower RMSE values overall and day 10 appears to have higher RMSE values. Weather data are fairly uniform (Appendix A) across days. It is not clear whether there is an environmental or activity explanation for these day-to-day differences in RMSE (the study's logbooks were not available to address this), or whether the differences are random. RMSE was not sensitive to uniform type; the RMSE mean and standard deviation for each uniform type were $0.28 \pm 0.024^{\circ}$C.

The accuracy of PSI based on the core temperature algorithm was also assessed for the field data. The LoA of ±1.2 is comparable to the variations that result from computing PSI with *CT* from a core temperature capsule. In the PSI scatter plot (Figure 7a), the line of identity and least squares regression line are well aligned with the line of identity, unlike the misalignment in the *CT* data in Figure 3(a). This is because heart rate forms half of the PSI score, and the same heart rate values are used for both observed and estimated PSI.

This page intentionally left blank.

6. LIMITATIONS

The two primary limitations in this study are: 1) the paucity of field data with core temperatures above 39°C, and 2) a study population that was limited to young men. An additional limitation is the lack of controlled data to assess elevated heart rate from a variety of factors other than activity and thermoregulation that can affect heart rate. These factors include diet, caffeine, sleep, and psychological stress.

Of the 47,549 data points that were used in the assessment, only 136 (0.3%) data points of the observed core temperature readings were above 39°C. Unfortunately, these are the most important temperatures to assess for applications related to heat injury. This is a common limitation across many field studies, particularly those studies that do not allow participation when observed core temperature exceeds 39.5°C.

The limitation to young men is also common across many field studies. No women were included in the database used to develop the algorithm and the developer has assessed the algorithm on only three women. A study concluded that gender differences result in no significant difference for PSI for the same levels of aerobic fitness [10], but the need to assess gender differences for the core temperature estimation algorithm remains. There is a similar need to evaluate older and less fit populations.

This page intentionally left blank.

7. CONCLUSION

The accuracy of a core temperature estimation algorithm that is based on a single parameter, heart rate, has been independently assessed. The assessment included both parametric variations and field data that were not used in algorithm development. Assessment through parametric variation showed qualitatively expected behavior. The field data set from a single study represents activities relevant to the military performed in a subtropical climate. The large number of data points (nearly 48,000) is comparable to the total number of data points evaluated by the algorithm developer, which drew from over 10 studies.

Overall, within the applicability of field data set that was evaluated, the core temperature estimation algorithm is sufficiently accurate and precise to provide a field-expedient indication of core temperature and thermal-work strain. Moreover, the algorithm accuracy is comparable to variations in temperature from an ingested temperature capsule. The algorithm is based on a data-driven model. As more data becomes available, the algorithm is expected to become more accurate over a wider range of conditions. Temperature estimation accuracy will also depend on the accuracy of heart rate measurements. Heart rate monitors that produce heart rates less accurate than those derived from post-processing the Hidalgo data will result in less accurate temperature estimates. An assessment of the accuracy of consumer-grade heart rate monitors is being completed as a separate study.

A follow-on independent assessment of the algorithm is recommended as additional field data become available to address the two primary limitations of this study: insufficient data with core body temperatures greater than 39°C and no data from women.

Finally, this report focuses on evaluating the algorithm for accuracy in estimating core temperature and physiological strain index. Quantifying the sensitivity and specificity of these measures for predicting heat injury remains for future work.

In conclusion, the core temperature estimation algorithm has been independently validated to be sufficiently accurate and precise to provide a field-expedient indication of core temperature and thermal-work strain, for young men with core temperatures up to 39^{o}C. More data with core temperatures greater than 39^{o}C and for women are needed for independent validation under those conditions, which will require a future data collection study.

This page intentionally left blank.

8. REFERENCES

1. M.J. Buller, W.J. Tharion, S.N. Cheuvront, S.J. Montain, R.W. Kenefick, J. Castellani, W.A. Latzka, W.S. Roberts, M. Richter, O.C. Jenkins and R.W. Hoyt, "Estimation of human core temperature from sequential heart rate observations," *Physiological Measurement,* pp. 781–798, 24 May 2013.

2. D. Casa, L. Armstrong, G. Kenny, F. O'Connor and R. Huggins, "Exertional Heat Stroke: New Concepts Regarding Cause and Care," *Current Sports Medicine Reports,* pp. 115–123, 2012.

3. B. Ely, M. Ely, S. Cheuvront, R. Kenefick, D. DeGroot and S. Montain, "Evidence against a 40 °C core temperature threshold for fatigue in humans," *J Appl Physiology,* vol. 107, pp. 1519–1525, 2009.

4. J.W. Domitrovich, J.S. Cuddy and B.C. Ruby, "Core-Temperature Sensor Ingestion Timing and Measurement Variability," *Jounral of Athletic Training,* vol. 45, no. 6, pp. 594–600, 2010.

5. D.S. Moran, A. Shitzer and K.B. Pandolf, "A physiological strain index to evaluate heat stress," *American Journal of Physiology,* 1 July 1998.

6. P.O. Astrand, "Aerobic work capacity in men and women with special reference to age," *Acta Physiologica,* vol. 49, no. 169, pp. 1–92, 1960.

7. M.J. Buller, W.J. Tharion, C.M. Duhamel and M. Yokota, "Real-time core body temperature estimation from heart rate for first responders wearing different levels of personal protective equipment," *Ergonomics,* 2015.

8. J.M. Bland and D.G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurements," *Lancet,* pp. 307–310, 1986.

9. J.S. Krouwer, "Why Bland–Altman plots should use *X*, not (*Y*+*X*)/2 when *X* is a reference method," *Statistics in Medicine,* pp. 27:778–780, 2008.

10. D. Moran, Y. Shapiro, A. Laor, S. Izraeli and K. Pandolf, "Can gender differences during exercise-heat stress be assessed by the physiological strain index?," *Am J Physiology,* vol. 276, pp. 1798–1804, 1999.

This page intentionally left blank.

APPENDIX A.
WEATHER DATA

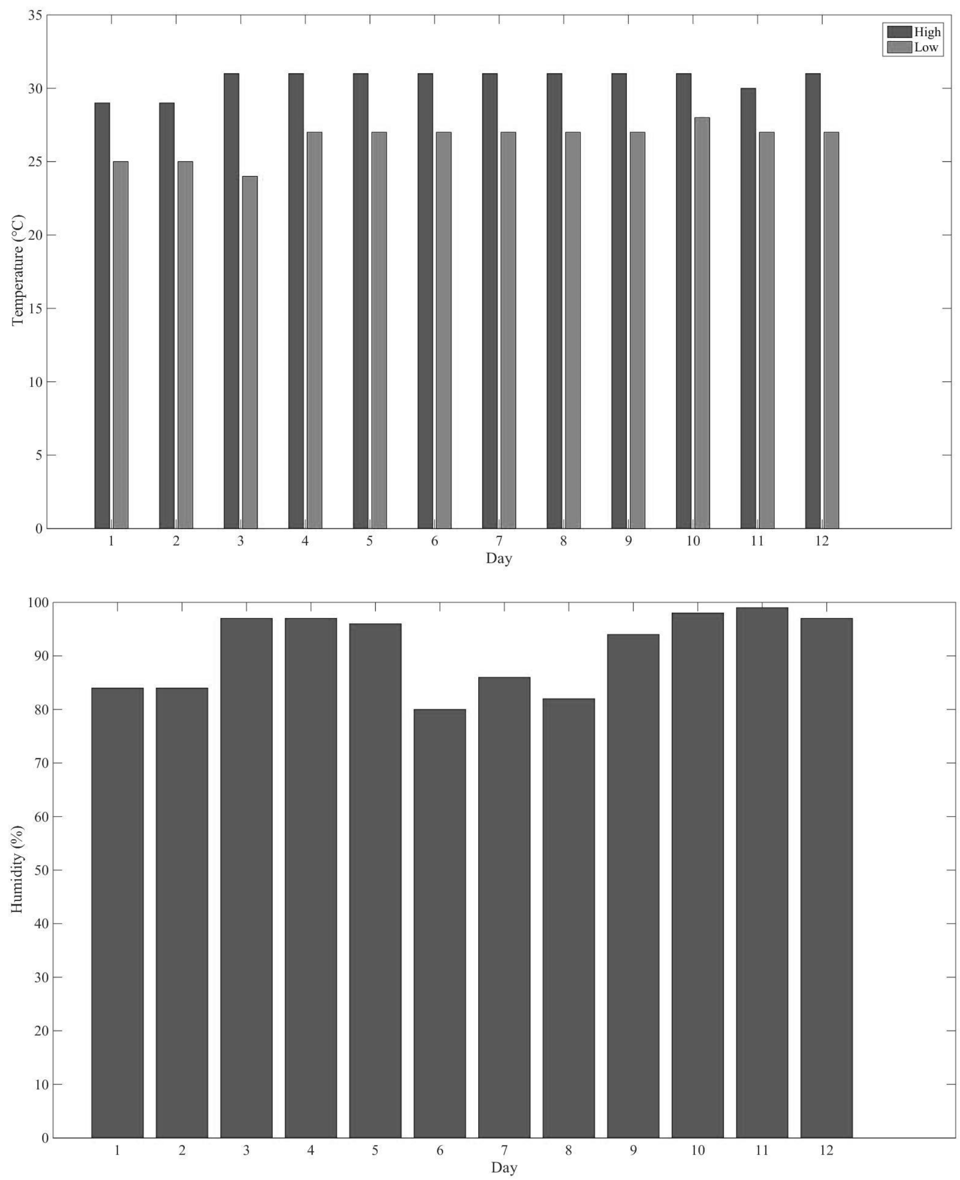


Figure A-1. (a) High and low temperatures for each study day, (b) Average humidity.

This page intentionally left blank.

APPENDIX B.
BREAKDOWN OF SELECTED DATA

Minutes Collected	Day												total (minutes)
Subject	1	2	3	4	5	6	7	8	9	10	11	12	
1	0	133	0	117	0	111	0	193	0	144	204	0	902
2	0	421	295	168	255	285	388	238	0	137	135	0	2322
3	278	0	291	0	353	370	392	224	319	183	0	0	2410
4	0	2	0	49	324	391	414	222	260	152	260	195	2269
5	277	56	0	63	179	72	0	0	0	0	0	0	647
6	0	333	0	138	0	310	349	65	200	32	0	0	1427
7	0	305	0	127	0	340	0	264	0	138	387	0	1561
8	0	271	0	160	348	199	235	0	0	132	414	215	1974
9	235	9	0	0	226	0	182	274	183	23	377	0	1509
10	0	283	302	0	0	345	399	265	0	183	39	0	1816
11	273	289	236	0	186	352	73	0	0	76	213	0	1698
12	0	264	0	184	0	250	0	148	0	167	344	82	1439
13	0	0	0	46	365	293	346	0	306	58	305	192	1911
14	395	445	292	126	0	0	91	157	136	55	163	200	2060
15	0	289	0	161	0	132	0	149	0	118	0	0	849
16	202	0	0	21	61	0	0	51	0	0	190	26	551
17	0	165	0	0	335	102	204	255	319	68	423	0	1871
18	0	331	166	0	0	128	0	123	0	0	2	0	750
19	345	441	0	152	179	263	260	0	0	137	413	24	2214
20	151	155	0	0	291	181	271	159	140	106	0	216	1670
21	0	446	0	161	0	353	0	0	0	172	0	0	1132
22	157	256	243	0	261	188	323	63	217	0	0	0	1708
23	212	371	0	131	322	293	0	0	0	65	0	0	1394
24	306	236	138	152	0	0	0	0	0	0	389	209	1430
25	0	0	0	0	0	83	0	0	0	0	0	0	83
26	172	456	0	148	0	0	394	59	0	46	369	0	1644
27	234	0	0	151	0	119	342	0	0	153	300	207	1506
28	250	435	0	143	174	0	0	0	0	0	192	0	1194
29	288	36	263	117	115	0	0	0	0	0	0	0	819
30	323	282	0	105	0	186	0	0	288	143	384	116	1827
31	0	361	0	179	278	241	373	0	0	114	370	119	2035
32	0	0	0	0	240	0	0	0	0	0	0	0	240
33	0	0	0	0	215	151	166	0	0	155	0	0	687
Total (minutes)	4098	7071	2226	2799	4707	5738	5202	2909	2368	2757	5873	1801	47549

Figure B-1. Duration of data for each subject and day that were used in the analysis. Green indicates greater than 2 hours of data collected; yellow, 0–2 hours; red, 0 data points. Blue highlights subject and day high and low.

This page intentionally left blank.

APPENDIX C.
BREAKDOWN OF RMSE VALUES

RMSE value	Day											
Subject	1	2	3	4	5	6	7	8	9	10	11	12
1	-	0.25	-	0.19	-	0.34	-	0.26	-	0.27	0.46	-
2	-	0.31	0.12	0.16	0.23	0.28	0.19	0.21	-	0.32	0.19	-
3	0.14	-	0.19	-	0.22	0.25	0.19	0.34	0.35	0.23	-	-
4	-	0.07	-	0.53	0.19	0.72	0.30	0.40	0.16	0.18	0.20	0.09
5	0.24	0.34	-	0.41	0.22	0.27	-	-	-	-	-	-
6	-	0.19	-	0.22	-	0.25	0.33	0.32	0.19	0.52	-	-
7	-	0.28	-	0.32	-	0.25	-	0.17	-	0.24	0.31	-
8	-	0.30	-	0.22	0.12	0.21	0.22	-	-	0.79	0.20	0.31
9	0.22	0.41	-	-	0.18	-	0.51	0.22	0.16	0.31	0.18	-
10	-	0.40	0.22	-	-	0.24	0.26	0.33	-	0.24	0.23	-
11	0.25	0.26	0.33	-	0.34	0.20	0.10	-	-	0.71	0.23	-
12	-	0.30	-	0.14	-	0.15	-	0.34	-	0.32	0.24	0.15
13	-	-	-	0.45	0.44	0.24	0.17	-	0.36	0.22	0.20	0.36
14	0.25	0.24	0.13	0.30	-	-	0.19	0.34	0.18	0.45	0.21	0.22
15	-	0.16	-	0.26	-	0.23	-	0.28	-	0.69	-	-
16	0.26	-	-	0.10	0.12	-	-	0.33	-	-	0.15	0.24
17	-	0.30	-	-	0.17	0.26	0.26	0.55	0.30	0.26	0.27	-
18	-	0.24	0.16	-	-	0.11	-	0.07	-	-	0.62	-
19	0.31	0.15	-	0.49	0.09	0.27	0.37	-	-	0.24	0.25	0.10
20	0.22	0.20	-	-	0.21	0.14	0.22	0.12	0.10	0.71	-	0.40
21	-	0.23	-	0.42	-	0.25	-	-	-	0.21	-	-
22	0.20	0.35	0.35	-	0.22	0.16	0.25	0.66	0.34	-	-	-
23	0.28	0.14	-	0.16	0.10	0.22	-	-	-	0.47	-	-
24	0.39	0.42	0.39	0.26	-	-	-	-	-	-	0.21	0.28
25	-	-	-	-	-	0.70	-	-	-	-	-	-
26	0.33	0.28	-	0.31	-	-	0.27	0.24	-	1.08	0.36	-
27	0.27	-	-	0.16	-	0.16	0.22	-	-	0.21	0.22	0.28
28	0.25	0.32	-	0.38	0.25	-	-	-	-	-	0.16	-
29	0.29	0.52	0.12	0.15	0.38	-	-	-	-	-	-	-
30	0.34	0.42	-	0.50	-	0.22	-	-	0.11	0.32	0.16	0.19
31	-	0.41	-	0.48	0.26	0.16	0.67	-	-	0.40	0.48	0.44
32	-	-	-	-	0.34	-	-	-	-	-	-	-
33	-	-	-	-	0.13	0.31	0.22	-	-	0.24	-	-

Figure C-1. RMSE value for each subject and day. Green coloring indicates lowest 10% of RMSE values. Red indicates highest 10% of RMSE values.

This page intentionally left blank.

APPENDIX D.
CORE TEMPERATURE ESTIMATION ALGORITHM

The following Matlab code implements the algorithm in [1] and was used to test the algorithm.

```
function [CT_out, v_out] = KFModel(HR, CTstart, v)
%KFMODEL estimate core temperature from heart rate with Kalman filter
% This version supports both batch mode (operate on entire HR time series)
% and incremental mode (update HR based on core temperature and confidence
% values from the previous time step).
%
% Usage examples:
%
% 1) batch mode (HR, v, and CT are vectors)
%       [CT, v] = KFModel(HR, 37.1, 0);
%       CT = KFModel(HR); % equivalent, uses default for initial values
%
% 2) incremental mode (CT1, CT2, v1, v2, HR1, HR2 are scalars)
%       [CT1, v1] = KFModel(HR1, 37.1, 0);
%       % do something with CT1 (such as update HSI value) then wait for next
update
%       [CT2, v2] = KFModel(HR2, CT1, v1);
%       % do something with CT2 (such as update HSI value) then wait for next
update
%       ...
%
% Inputs:
% HR = In batch, vector of minute to minute HR values; in incremental, HR
value
%       at current timestep
% CTstart = Core Body Temperature at time 0 (scalar, default 37.1, or CT
%            at previous timestep if incremental)
% v = Confidence of start value (scalar, default 0, of v at previous timestep
%      if incremental
%
% Outputs:
% CT_out = In batch, vector of minute to minute CT estimates (same size as
HR);
%           in incremental, core temp estimate at current time step
% v_out = In batch, vector of minute to minute confidence values (same size
as HR);
%          in incremental, confidence value at current time step
%
% Reference:
% Buller, Mark J., et al. "Estimation of human core temperature from
sequential
% heart rate observations." Physiological measurement 34.7 (2013): 781.
%
% Contributors:
% - Mark Buller, developed original version of this function (batch only)
% - Jeff Simpson, modified function to operate in incremental or batch mode
% - Delsey Sherrill, added documentation and edited code for readability
%
```

```
%
% last updated

if nargin < 2
    CTstart = 37.1; % degrees Celsius
end
if nargin < 3
    v = 0;
end

%Extended Kalman Filter Parameters
a = 1; gamma = 0.022^2;
b_0 = -7887.1; b_1 = 384.4286; b_2 = -4.5714; sigma = 18.88^2;

%Initialize Kalman filter
x = CTstart;
CT_out = zeros(size(HR));
v_out = zeros(size(HR));

%Iterate through HR time sequence
for time = 1:length(HR)

    %Time Update Phase
    x_pred = a*x; %Equation 3
    v_pred = (a^2)*v+gamma; %Equation 4

    %Observation Update Phase
    z = HR(time);
    c_vc = 2.*b_2.*x_pred+b_1; %Equation 5
    k = (v_pred.*c_vc)./((c_vc.^2).*v_pred+sigma); %Equation 6
    x = x_pred+k.*(z-(b_2.*(x_pred.^2)+b_1.*x_pred+b_0)); %Equation 7
    v = (1-k.*c_vc).*v_pred; %Equation 8
    v_out(time) = v;
    CT_out(time) = x;
end

% x is the core temp at each step
% a doesn't change
% gamma doesn't change
% v changes!!
% z is HR
```

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	TBD 2016
2. REPORT TYPE	Project Report
3. DATES COVERED *(From - To)*	
4. TITLE AND SUBTITLE	Validation of Core Temperature Estimation Algorithm
5a. CONTRACT NUMBER	FA8721-05-C-0002
5b. GRANT NUMBER	
5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)	K.M. Showers, A.R. Hess, and B.A. Telfer, MIT Lincoln Laboratory
5d. PROJECT NUMBER	2489
5e. TASK NUMBER	1
5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02420-9108
8. PERFORMING ORGANIZATION REPORT NUMBER	PSM-4
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	U.S. Army Medical Materiel Development Activity 1430 Veterans Drive Fort Detrick, MD 21702-5001
10. SPONSOR/MONITOR'S ACRONYM(S)	USAMMDA
11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT	Approved for public release; distribution is unlimited.
13. SUPPLEMENTARY NOTES	

14. ABSTRACT

Real-time knowledge of core body temperature is a key input to determine thermal work-strain and risk of heat injury. An algorithm for estimating core temperature based on heart rate has been developed by others in order to avoid standard but more invasive measurements, such as ingestible capsules and rectal or esophageal probes. This report provides an independent assessment of the algorithm, based on both parametric variations and field data that were not used in algorithm development. Assessment through parametric variation shows qualitatively expected behavior. Field data were taken from a study of 33 young, male, military personnel who tested six prototype uniforms over the course of 12 days in Okinawa, Japan. The field data selected for assessment consist of nearly 48,000 measurements of heart rate and core temperature. Core temperature was measured by an ingestible capsule. The observed core temperature range was 36.1–39.5°C. Bland-Altman analysis yielded a bias of –0.01°C and a 95% Limit of Agreement (LoA) of ±0.58°C. The LoA is similar to the ±0.5°C LoA resulting from variations in ingestible capsule temperatures, and is also consistent with the ±0.63°C LoA found by the algorithm developer on other field data. The accuracy of the Physiological Strain Index (PSI), a standard estimate of thermal-work strain on a nominal 0–10 scale, was also assessed with the estimated core temperature as an input. A bias of –0.01 and LoA of ±1.2 was found. The PSI LoA is comparable to the LoA resulting from the variability in computing PSI with two different ingestible capsule temperatures. Overall, within the applicability of the field data that were evaluated, the core temperature estimation algorithm is sufficiently accurate and precise to provide a field-expedient indication of core temperature and thermal-work strain. Follow-on independent assessments of the algorithm are recommended as additional field data become available for higher core body temperatures and from women. This report focuses on evaluating the algorithm for accuracy in estimating core temperature and PSI. Quantifying the sensitivity and specificity of these measures for predicting heat injury remains for future work.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	Same as report	36	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18