
Copyright:
© 2014 the American Physiological Society Licensed under Creative Commons Attribution CC-BY 3.0

DOI link to article:
http://dx.doi.org/10.1152/japplphysiol.00421.2014

Date deposited:
05/08/2015

This work is licensed under a Creative Commons Attribution 3.0 Unported License
Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents

Vincent T. van Hees,1,7 Zhou Fang,2,3 Joss Langford,3 Felix Assah,4 Anwar Mohammad,5 Inacio C. M. da Silva,6,7 Michael I. Trenell,1 Tom White,7 Nicholas J. Wareham,7 and Søren Brage7
1MoveLab, Institute of Cellular Medicine, Newcastle University, Newcastle, United Kingdom; 2Department of Statistics, University of Oxford, Oxford, United Kingdom; 3Activinsight, Limited, Kimbolton, United Kingdom; 4University of Yaoundé, Yaoundé, Cameroon; 5Dasmus Diabetes Institute, Kuwait City, Kuwait; 6Federal University of Pelotas—Postgraduate Program in Epidemiology, Pelotas, Brazil; and 7Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

Submitted 20 May 2014; accepted in final form 3 August 2014

van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva IC, Trenell MI, White T, Wareham NJ, Brage S. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. J Appl Physiol 117: 738–744, 2014. First published August 7, 2014; doi:10.1152/japplphysiol.00421.2014.—Wearable acceleration sensors are increasingly used for the assessment of free-living physical activity. Acceleration sensor calibration is a potential source of error. This study aims to describe and evaluate an autocalibration method to minimize calibration error using segments within the free-living records (no extra experiments needed). The autocalibration method entailed the extraction of nonmovement periods in the data, for which the measured vector magnitude should ideally be the gravitational acceleration (1 g); this property was used to derive calibration correction factors using an iterative closest-point fitting process. The reduction in calibration error was evaluated in data from four cohorts: UK (n = 921), Kuwait (n = 120), Cameroon (n = 311), and Brazil (n = 200). Our method significantly reduced calibration error in all cohorts (P < 0.01), ranging from 16.6 to 3.0 mg in the Kuwaiti cohort to 76.7 to 8.0 mg error in the Brazil cohort. Utilizing temperature sensor data resulted in a small nonsignificant additional improvement (P > 0.05). Temperature correction coefficients were highest for the z-axis, e.g., 19.6-mg offset per 5°C. Further, application of the autocalibration method had a significant impact on typical metrics used for describing human physical activity, e.g., in Brazil average wrist acceleration was 0.2 to 51% lower than uncalibrated values depending on metric selection (P < 0.01). The autocalibration method as presented helps reduce the calibration error in wearable acceleration sensor data and improves comparability of physical activity measures across study locations. Temperature utilization seems essential when temperature deviates substantially from the average temperature in the record but not for multiday summary measures.

calibration; accelerometer; physical activity; epidemiology; GENE-Activ

WEARABLE ACCELEROMETERS ARE increasingly used in the assessment of physical activity (2, 4, 6). In recent years accelerometers have become available that are feasible for long-term monitoring of behavior in population studies, while at the same time being capable of storing weeklong data in g-units (1 standard g = 9.80665 m/s²) at a sample frequency high enough to capture the main frequencies of body movement, referred to as raw data accelerometry (12). Population studies collecting raw accelerometer data include surveillance studies like NHANES (26) in the U.S. and national biobanks such as UK Biobank (27).

An acceleration sensor works on the principle that acceleration is captured mechanically and converted into an electrical signal, which depending on the sensor type is either a voltage, a resistance, or a capacitance (13). The relationship between the electrical signal and the acceleration is usually assumed to be linear, involving an offset and a gain factor. We shall refer to the establishment of the offset and gain factor as the sensor calibration procedure (5, 18). Accelerometers are usually calibrated as part of the manufacturing process under nonmovement conditions using the local gravitational acceleration as a reference (5, 18). The manufacturer calibration can later be evaluated by holding each sensor axis parallel (up and down) or perpendicular to the direction of gravity; readings for each axis should be ±1 and 0 g, respectively (5, 18).

However, this procedure can be cumbersome in studies with a high throughput. Furthermore, such a calibration check will not be possible for data that have been collected in the past and for which the corresponding accelerometer device does not exist anymore. Techniques have been proposed that can check and correct for calibration error based on the collected triaxial accelerometer data in the participant’s daily life without additional experiments, referred to as autocalibration (6a, 8–10, 19). The general principle of these techniques is that a recording of acceleration is screened for nonmovement periods. Next, the moving average over the nonmovement periods is taken from each of the three orthogonal sensor axes and used to generate a three-dimensional ellipsoid representation that should ideally be a sphere with radius 1 g, see example in Fig. 1. Here, deviations between the radius of the three-dimensional ellipsoid and 1 g (ideal calibration) can then be used to derive correction factors for sensor axis-specific calibration error (6a, 8–10, 19).

Previously published work on autocalibration techniques focused on the technical description and proof of concept but did not demonstrate feasibility and accuracy in wrist accelerometer data collected under real study conditions, involving participants under free-living conditions (daily life) and in a diverse sample of the global population (6a, 8–10, 19). Fur-
thermore, it remains unknown whether autocalibration has a significant impact on acceleration metrics typically used for physical activity assessment.

Temperature has been identified as a potential source of calibration error in low cost acceleration sensors (20). The specification sheet of the acceleration sensor chip used in the GENEActiv accelerometer (ADXL345; Analog Devices) as used in this study, indicates that a change of 1°C relative to 25°C could result in a 0.4 to 1.2 m\(^g\) (1 g = 1,000 m\(^g\)) change in acceleration value (1). It could therefore be hypothesized that the availability of temperature information alongside measurement of acceleration may aid the autocalibration process.

The current study aims to describe an autocalibration method that can be configured to take into account a potential temperature dependency of the sensor’s response to acceleration. The second aim is to implement and evaluate the autocalibration method in a diverse sample of the global population. The third and final aim is to demonstrate the degree to which application of the autocalibration method has any significant impact on metrics derived for physical activity assessment.

METHODS

Population data. The autocalibration method was evaluated based on data collected with wrist-worn raw accelerometry in subsamples of epidemiological cohorts from Africa, Europe, South America, and the Middle-East, representing locations with different gravity. Cohorts included: The Fenland Study (Cambridgeshire, UK) (21), a repeated cross-sectional survey of the Cameroon Physical Activity Study (3), the Kuwait Wellbeing Study, and the 1993 Pelotas birth cohort (Brazil) (28). The data subsamples in each cohort span most local seasons and represent very diverse populations, lifestyles, and environmental conditions. Basic cohort characteristics are described in Table 1.

The same accelerometer brand was used in all cohorts (GENEActiv; Activinsights, Kimbolton, UK). This accelerometer includes a triaxial acceleration sensor (ADXL345) with a ±8-g dynamic range and a 12-bit resolution and a temperature sensor (MCP9700T). Most of the devices used in the UK and Brazil cohort were older (lower serial number) than the devices used in the Kuwait and Cameroon cohorts. In all cohorts, participants were asked to wear the accelerometer on their nondominant wrist during sleeping and waking hour. All participants provided informed consent, and each study was approved by the local ethics committee.

Autocalibration method. Two versions of the autocalibration method were designed and evaluated; one based on acceleration data only (C1) and one based on both acceleration and temperature (C2). For every consecutive time window of 10 s in a particular data record, the following signal features were extracted: average acceleration per axis, standard deviations in the acceleration per axis, and average temperature. For the calibration procedure, only time windows for which the standard deviation was $<13$ mg in all three axes were retained. Here, 13 mg was selected just above the empirically derived baseline (noise) standard deviation of 10 mg to retain only nonmovement periods. The resulting set of time windows, or calibration epochs, for each of the three axes can be presented in a three-dimensional space as an ellipsoid (22), an example of which is shown in Fig. 1. The deviation between 1 g and the Euclidean norm ($\sqrt{a_x^2 + a_y^2 + a_z^2}$) of the acceleration of the three axes is an indication of calibration error. Next, the axis-specific calibration for $C_1$ can be defined as: $s_i(t) = d_i + s_i(t) - a_i$. Here, $s_i(t)$ and $s_i(t)$ correspond to the acceleration signal before and after correction, respectively, $i$ is the
The formula is:

\[
\text{g.calibrate} = \hat{\text{ai}}(\text{t}) + \text{si(}\text{t})\text{ai} + [\text{T}(\text{t}) - \text{c}m]\text{.,}
\]

Here, \(T(\text{t})\) is the temperature at time point \(\text{t}\), \(\text{c}\) is the average temperature in the ellipsoidal data as used for the autocalibration procedure, and \(m\) is the axis specific temperature-related offset corrections factor. The average temperature acts like a fixed reference point relative to which axis specific temperature-related offset corrections factor. The average values to a sphere (C1) or a hypercylinder (C2) was used to fit the impact of data resolution boundaries. Therefore, a calibration error reduction to <10 mg was considered acceptable. If the calibration error after autocalibration was higher than before autocalibration, then correction factors were replaced by default values 1 and 0 for gain and offset, respectively. The latter was done to avoid a negative influence of autocalibration on the data.

The method has been released as function \(\text{g.calibrate}\) in R-package \(\text{GGIR}\), which currently works with binary data collected with the accelerometer used in the current study as well as its predecessor, \(\text{GENEA}\) (14). Additionally, an extract of the R-code related to the ICP fitting process is provided in the APPENDIX.

**Evaluation.** The absolute difference between 1 g and the Euclidean norm of the values of the three axes was averaged per measurement file (1 file = 1 participant) and used as an indicator of calibration error before autocalibration (C0), following autocalibration without temperature compensation (C1), and following autocalibration with temperature compensation (C2).

Further, we assessed the impact of autocalibration on population estimates of physical activity using two commonly used metrics of body movement: the Euclidean Norm Minus One with negative values rounded up to zero (ENMO) and band-pass filtering of three axis accelerometer used in the current study as well as its predecessor, \(\text{GENEA}\) (14). Additionally, an extract of the R-code related to the ICP fitting process is provided in the APPENDIX.

**Evaluation.** The absolute difference between 1 g and the Euclidean norm of the values of the three axes was averaged per measurement file (1 file = 1 participant) and used as an indicator of calibration error before autocalibration (C0), following autocalibration without temperature compensation (C1), and following autocalibration with temperature compensation (C2).

Further, we assessed the impact of autocalibration on population estimates of physical activity using two commonly used metrics of body movement: the Euclidean Norm Minus One with negative values rounded up to zero (ENMO) and band-pass filtering of three axis accelerometer used in the current study as well as its predecessor, \(\text{GENEA}\) (14). Additionally, an extract of the R-code related to the ICP fitting process is provided in the APPENDIX.

**Evaluation.** The absolute difference between 1 g and the Euclidean norm of the values of the three axes was averaged per measurement file (1 file = 1 participant) and used as an indicator of calibration error before autocalibration (C0), following autocalibration without temperature compensation (C1), and following autocalibration with temperature compensation (C2).

Finally, to estimate the relative importance of correcting offset or gain factors we selected a random sample of 20 accelerometer record-

---

**Table 1. Cohort characteristics**

<table>
<thead>
<tr>
<th>Cohort</th>
<th>UK</th>
<th>Kuwait</th>
<th>Cameroon</th>
<th>Brazil</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n) (male/female)</td>
<td>407/514</td>
<td>72/48</td>
<td>144/167</td>
<td>100/100</td>
</tr>
<tr>
<td>Age, yr</td>
<td>50.3 (7.2)</td>
<td>43.0 (10.7)</td>
<td>40.3 (12.6)</td>
<td>18.4 (19–19)†</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>77.1 (16.1)</td>
<td>81.8 (18.3)</td>
<td>76.8 (15.2)</td>
<td>65.8 (14.7)</td>
</tr>
<tr>
<td>Height, m</td>
<td>170.0 (9.6)</td>
<td>167.5 (8.5)</td>
<td>166.7 (8.4)</td>
<td>167.3 (8.3)</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>26.5 (4.6)</td>
<td>29.0 (5.3)</td>
<td>28.2 (9.6)</td>
<td>23.4 (4.7)</td>
</tr>
<tr>
<td>Monitor protocol, days</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Sample frequency, Hz</td>
<td>60</td>
<td>50</td>
<td>100</td>
<td>85.7</td>
</tr>
<tr>
<td>Geographic latitude, °</td>
<td>52.2 N</td>
<td>29.4 N</td>
<td>51.1 N</td>
<td>31.8 S</td>
</tr>
<tr>
<td>Altitude, m</td>
<td>6</td>
<td>20</td>
<td>7261 (600)</td>
<td>7</td>
</tr>
<tr>
<td>Magnitude of gravity, m-s⁻²</td>
<td>9.8127</td>
<td>9.7928</td>
<td>9.7807</td>
<td>9.7947</td>
</tr>
<tr>
<td>Difference in gravity relative to UK, mg</td>
<td>0.0</td>
<td>−2.0</td>
<td>−3.3</td>
<td>−1.8</td>
</tr>
<tr>
<td>Seasonal distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Dec-Feb</td>
<td>23%</td>
<td>32%</td>
<td>47%</td>
<td>39%</td>
</tr>
<tr>
<td>In Mar-Apr</td>
<td>24%</td>
<td>27%</td>
<td>0%</td>
<td>10%</td>
</tr>
<tr>
<td>In Jun-Aug</td>
<td>30%</td>
<td>41%</td>
<td>13%</td>
<td>0%</td>
</tr>
<tr>
<td>In Sep-Nov</td>
<td>23%</td>
<td>1%</td>
<td>41%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Data are expressed as mean (SD). BMI, body mass index. *According to calculation with World Geodetic System 1984; †age range; ‡Yaounde and Bamenda.
between autocalibration with temperature utilization (C2), calibration error was observed in any of the four cohorts (see Table 3). However, no significant further reduction in performance is affected when optimizing only offset or gain, with the corresponding other set of factors fixed to 1 (gain) or 0 (offset), respectively.

**Statistics.** All statistical analyses were conducted in R (http://cran.r-project.org/). Wilk’s lambda test was used to compare the three autocalibration configurations across all percentiles. If Wilk’s lambda test indicated a significant difference, then repeated measures ANOVA was used to compare the three autocalibration configurations per metric, using the function lme from the nlme-package and the function anova from the stats-package (20a). Post hoc pair-wise Tukey tests were performed using the function glht from the multcomp package (16). Significance was set at P < 0.05.

### RESULTS

Average calibration correction factors are reported in Table 2. Application of the autocalibration method significantly reduced calibration error in all cohorts (P < 0.01), with improvements being greatest in the Brazilian cohort (from 76.7 to 8.0 mg) and smallest in the Kuwaiti cohort (from 16.6 to 3.0 mg; see Table 3). However, no significant further reduction in calibration error was observed in any of the four cohorts between autocalibration with temperature utilization (C2), compared with that without temperature utilization (C1) (P > 0.05; see Table 3). The percentage of files with calibration error under 10 mg was 6.1, 94.4, and 99.0% for C0, C1, and C2 respectively (Pearson’s Chi-squared: $\chi^2 = 3,816.0$, df = 2, $P < 0.0001$). An animation of the calibration ellipsoid before and after calibration can be found on our website: http://www.mrc-epid.cam.ac.uk/research/resources.

Application of the autocalibration method had a significant impact on the average and distribution of acceleration metric output in each of the four cohorts ($F > 5.8$, $P < 0.001$). The magnitude of the difference between C0 and C1 for metric $BFEN$ was systematically $< 1$ mg, which was in contrast to metric $ENMO$ for which differences of 20 mg and higher were observed between C0 and C1 (see Tables 4 and 5). Post hoc Tukey analyses revealed no significant difference in metric output between C1 and C2, except for the lower range in the distribution of acceleration values in the UK cohort, see Tables 4 and 5.

The minimum within-person temperature range observed within the ellipsoid data was 8.8, 9.7, 6.4, and 7.1°C for UK, Kuwait, Cameroon, and Brazil, respectively.

For the evaluation of the relative importance of offset and gain (Pelotas subset), autocalibration based on only offset correction or only gain correction reduced a 65.0 ± 26.7 mg
respectively, while optimizing both offset and gain resulted in

$P_k$ averages; Brazil

Cameroon

of acceleration of 0.0196 given a temperature difference of 5°C would result in a change
correction factor for the study conditions. Temperature utilization did not result in a
difference in the precision gain of autocalibration between UK and Brazil on the one hand and Cameroon and Kuwait on the other hand may indicate that the relatively newer devices used for Cameroon and Kuwait have less calibration error. Again, a lack of standardized conditions complicates this comparison. It is also important to note that the proposed method effectively expresses all data relative to local gravity that has known geographical variation; one would need to multiply with the magnitude of local gravity to convert to absolute acceleration in meters per second squared. Despite the challenges in directly comparing the four cohorts, the results stratified by cohort illustrate that the method succeeds in reducing error in each of the four study settings and with an impact on typical physical activity summary measures proportional to baseline calibration error ($C_0$).

discussion

The autocalibration method as presented allows for a significant reduction in average calibration error under a wide range of study conditions. Temperature utilization did not result in a significant further reduction of average calibration error for the measures selected. However, inspection of the derived temperature offset correction factors (Table 2) indicates that temperature utilization could be essential for sections of the signal with temperature conditions far away from the average temperature. For example, in the UK cohort the average temperature offset correction factor for the $z$-axis was 0.00392 (Table 2), which given a temperature difference of 5°C would result in a change of acceleration of 0.0196 g ($5 \times 0.00392$ g). A value of 19.6 mg may be considered high in the context of the acceleration

value distribution as provided in Tables 4 and 5. The significant difference as found between $C_1$ and $C_2$ in the lower end of the metric value distribution in the UK cohort hints at an impact of temperature utilization that will only be visible in the most inactive parts of a day. Considering that sleep is likely to take up >25% of a day, it seems unlikely that the temperature dependency of the 5th and 25th percentiles relates to wake-time behavior. Instead, accounting for temperature dependency may help to improve estimates of monitor nonwear time and the detection of sleep stages in future research.

The implementation of the autocalibration method had a significant impact on the average and distribution of metric outputs; however, substantially more so for metric ENMO compared with metric BFEN. In our previous study we observed that metrics ENMO and BFEN are highly correlated but not identical (11). Metric ENMO may be more appropriate for energy expenditure estimation and easier for researchers to describe, replicate, and interpret (11). In addition, the frequency filtering as part of metric BFEN effectively reduces calibration offset error, which explains why the autocalibration procedure as evaluated here shows only minor impact on these estimates. Temperature changes tend to be slow, which the band-pass filter would catch and remove as low frequency components (11). We conclude from this that autocalibration will have an important impact on studies that rely on average and distribution characteristics of metric ENMO but much less so for metric BFEN. Note that these findings should not be confused for the validity of metric BFEN or ENMO.

The strong relative importance of offset correction as seen in the subsample of 20 individuals combined with the fairly constant absolute difference between the cohort percentiles corresponding to $C_0$ and $C_1$ (Tables 4 and 5) indicates that the offset calibration has a bigger impact compared with gain calibration. Translating this observation to physical activity research means that the impact of calibration error and therefore the benefit of autocalibration will be relatively high for physical activities involving low acceleration and relatively low for activities involving high magnitude accelerations.

Results indicate that the autocalibration method works under a wide range of experimental conditions, spanning different geographical latitudes, different seasons affecting temperature variation during the day, different populations affecting movement and activity patterns, different built environments, and different adult age groups. Nonetheless, the dataset as presented is insufficient to investigate the causal relationship between specific study conditions and calibration error. The difference in the precision gain of autocalibration between UK and Brazil on the one hand and Cameroon and Kuwait on the other hand may indicate that the relatively newer devices used for Cameroon and Kuwait have less calibration error. Again, a lack of standardized conditions complicates this comparison. It is also important to note that the proposed method effectively expresses all data relative to local gravity that has known geographical variation; one would need to multiply with the magnitude of local gravity to convert to absolute acceleration in meters per second squared. Despite the challenges in directly comparing the four cohorts, the results stratified by cohort illustrate that the method succeeds in reducing error in each of the four study settings and with an impact on typical physical activity summary measures proportional to baseline calibration error ($C_0$).
The current study was done with wrist-worn accelerometers. Compared with other body locations wrist attachment may allow for easier collection of sparse ellipsoidal data and by that enhancing the autocalibration process. Therefore, caution is needed when implementing this method on data collected from other body locations.

In conclusion, the autocalibration method as presented reduces the calibration error in acceleration data from wrist-worn sensors as collected on four continents. Temperature utilization produces the calibration error in acceleration data from wrist-worn sensors as collected on four continents. Temperature utilization enhances the autocalibration process. Therefore, caution is needed when implementing this method on data collected from other body locations.

In conclusion, the autocalibration method as presented reduces the calibration error in acceleration data from wrist-worn sensors as collected on four continents. Temperature utilization enhances the autocalibration process. Therefore, caution is needed when implementing this method on data collected from other body locations.


click to continue

APPENDIX I: EXTRACT OF R-CODE RELATED TO ICP PROCEDURE FROM R-PACKAGE GGIR

The variable “input” is the average acceleration per axis per epoch provided as a matrix with three columns corresponding to the three axis. Variable “inputtemp” is the average temperature per epoch provided as a matrix with the temperature values replicated in three columns.

```r
meantemp = mean(as.numeric(inputtemp[,1]))
inputtemp = inputtemp - meantemp
translate = rep(0, ncol(input)); gain = rep(1, ncol(input))
tempposset = rep(0, ncol(input))
weights = rep(1, nrow(input))
res = Inf
maxter = 1000
tol = 1e-10
for (iter in 1:maxter) {
curr = gain(input, center = -translate, gain = 1/gain) +
inputtemp - meantemp; F = 1/tempposset
closestpoint = curr/sqrt(rowSums(curr^2))
k = 1
translatech = rep(0, ncol(input)); gainch = rep(1, ncol(input))
toffch = rep(0, ncol(input))
folb = lm.wfit(clipboard1, curr[k], inputtemp[k], closestpoint[k, drop = F], w = weights)
if (use.temp == TRUE) translatech[k] = folb$coeff[1]
if (use.gain == TRUE) gainch[k] = folb$coeff[2]
if (use.offset == TRUE) toffch[k] = folb$coeff[3]
curr[k] = folb$fitted.values
}
translate = translate + translatech / (gain * gainch)
if (use.temp == TRUE) tempposset = tempposset * gainch + toffch
gain = gain * gainch
res = c(res, 3 * mean(weights*(curr-closestpoint)^2)/sum(weights))
weights = pmin(1/sqrt(rowSums((curr - closestpoint)^2)), 1/0.01)
if (abs(res[iter+1] - res[iter]) < tol) break
```

ACKNOWLEDGMENTS

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had full responsibility for the decision to submit for publication.

GRANTS

This research was supported by funding from the Wellcome Trust, the Medical Research Council (MC_UU_12015/3), and His Highness Sheikh Nasser Al-Mohammad Al-Sabah, and Dasman Diabetes Institute.

DISCLOSURES

Joss Langford is employed by Activinsights We declare that we have no competing interests.

AUTHOR CONTRIBUTIONS


REFERENCES