Co-funded by the European Union

Ministry of the Environment of the Czech Republic

LERCO

Methodological approach

Detection of spatiotemporal parameters from biomechanical and Fitbit data

(2025)

Co-funded by the European Union

Ministry of the Environment of the Czech Republic

LERCO

DESCRIPTION OF THE METHOD

Title

Detection of spatiotemporal parameters from biomechanical and Fitbit data

Author list

Jan Šustek, jan.sustek@osu.cz, 1

Jan Urbaczka, jan.urbaczka@osu.cz, 2

Daniel Jandačka, daniel.jandacka@osu.cz, 2

Steriani Elavsky, steriani.elavsky@osu.cz 2

1 - Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, Czech Republic

2 - Department of Human Movement Studies, University of Ostrava, Czech Republic

Annotation

This methodical material provides a comprehensive framework for the computation and interpretation of walking and running spatiotemporal parameters derived from consumer-grade wearable technology data (Fitbit Charge 3,4 wristbands) supplemented by the walking and running biomechanical data from the 4HAIE study. The document outlines a methodological approach for sequencing, analysis, and computation of combined data from inertial and optical sensors, focusing on extracting running/walking metrics such as stride length, covered distance, and pace. Particular attention is paid to algorithmic preprocessing, noise filtering, signal segmentation, and validation strategies relative to gold-standard biomechanical and Global Position System data.

Purpose

In the context of sports biomechanics and applied human locomotion research, this work bridges the gap between laboratory-grade data fidelity and the ecological validity offered by consumer wearable technologies in free-living conditions. The methodology is especially pertinent for longitudinal field studies, large-cohort interventions, and athlete monitoring protocols where high-frequency data capture must be balanced with feasibility and participant compliance.

Leveraging the widespread availability of wrist-worn devices, this material addresses the opportunities and limitations inherent in translating accelerometric, gyroscopic, and barometer data into valid constructs for future data applications.

Method

Data Collection

The data from the 4HAIE cohort comprising Czech active runners and inactive controls aged 18 to 65 years were used (Jandacka et al., 2020). The raw Fitbit data in .JSON format were derived using the manufacturer's web application interface (API) and aggregated using the HealthReact platform (https://www.healthreact.eu/). Subsequently, these data were converted into .CSV format for further processing. The final data set included:

- **STEPS.csv** - Minute-resolution step count data.
- **DAILY_ACTIVITY_SUMMARY_ACTIVITY.csv** - Fitbit's automated activity summaries.
- **DISTANCE.csv** - Minute-level GPS-derived distance estimates.
- **ACTIVITY_LOG_LIST.csv** - Metadata describing activity source and type.
- **HR.csv** and **RESTING.csv** - Heart rate data and daily resting heart rate based on heart rate sensor.
- **ELEVATION.csv** - Elevation gain data based on barometer sensor data.
- **Biomechanical data**: Gait parameters data based on biomechanical measurements of walking and running (Jandacka et al., 2020).

All data were aggregated and matched using participant identifiers and timestamps.

Activity Detection

Based on the values in the file STEPS.csv, activity bouts were identified through a three-phase detection process involving sliding window analysis and minimum duration criteria. Detected segments were further parsed into walking and running subsegments, with false positives eliminated through cross-referencing against predefined activity categories from Fitbit's automated classification system (DAILY_ACTIVITY_SUMMARY_ACTIVITY.csv).

Day segmentation

When detecting activities, the data was first divided by days. Each activity thus falls into one day. This day in the format yyyy-mm-dd is stored in the "day" column. Any activity lasting over midnight was consequently counted as two shorter activities in two separate days. This also ensured a seamless inclusion in the weeks.

Week segmentation

The "week" column indicates the number of the week. A week is counted as seven consecutive days from Monday to Sunday. The week number 1 is the week from 2019-01-07. Week numbers are counted continuously in subsequent years; for example, the week from 2020-01-06 has the number 53.

Step thresholds and filtering

Based on large-scale observational studies, walking was defined as a cadence between 60 and 140 steps per minute, where 60 steps/min mark the lower bound for purposeful classification, below which activity likely reflects incidental movements or noise[1]. The transition zone between walking and running was 135–140 steps/min[2]. The threshold was adjusted in this range based on the data of the particular participant. The natural cadence aligning with running bouts in various experience-level runners was previously reported in a range of 141 to 205 steps/min[1,3]. Based on our data, cadences above 240 steps/min were not utilized for subsequent analysis to minimize misclassification from high-frequency arm movements and noise, as sustained step rates above this level are atypical and rarely represent authentic running[4].

The following threshold parameters were used:

- **Walking threshold (WALK_THRESHOLD)** = 60 steps/min
- **Running threshold (RUN_THRESHOLD)** = 140 steps/min
- **Maximum valid steps (MAX_STEPS)** = 240 steps/min
- **Minimum bout duration (MIN_DURATION)** = 8 min
- **Sliding window width (WINDOW_WIDTH)** = 5 min
- **Maximum bad values per window (WINDOW_TOLERANCE)** = 2 min

If any minute exceeded MAX_STEPS, the step count was replaced with zero (error correction).

---

[1] Tudor-Locke, C., Aguiar, E. J., Han, H., Ducharme, S. W., Schuna, J. M., Barreira, T. v., Moore, C. C., Busa, M. A., Lim, J., Sirard, J. R., Chipkin, S. R., & Staudenmayer, J. (2019). Walking cadence (steps/min) and intensity in 21-40 year olds: CADENCE-adults. International Journal of Behavioral Nutrition and Physical Activity, 16(1). https://doi.org/10.1186/s12966-019-0769-6

[2] Chase, C. J., Aguiar, E. J., Moore, C. C., Chipkin, S. R., Staudenmayer, J., Tudor-Locke, C., & Ducharme, S. W. (2023). Cadence (steps/min) as an indicator of the walk-to-run transition. Human Movement Science, 90. https://doi.org/10.1016/j.humov.2023.103117

[3] Gamez-Paya, J., Aladro-Gonzalvo, A. R., Marcos, D. G. de, Villarón-Casales, C., & Amo, J. L. L. del. (2023). Footstrike Pattern and Cadence of the Marathon Athletes at the Tokyo 2020 Olympic Games. Applied Sciences (Switzerland), 13(11). https://doi.org/10.3390/app13116620

[4] Han, H., Kim, H., Sun, W., Malaska, M., & Miller, B. (2020). Validation of wearable activity monitors for real-time cadence. Journal of Sports Sciences, 38(4), 383–389. https://doi.org/10.1080/02640414.2019.1702281

General Activity Bout Detection

Walking and running bouts were identified in three phases:

1. First, the segments where there was some activity suspected to be walking or running were identified.
   - Identified segments where ≥ WALK_THRESHOLD steps/min were maintained.
     - The beginning of the segment had the value ≥ WALK_THRESHOLD.
   - A sliding window was used to allow brief periods below the threshold without terminating a bout.
     - Periodically, it was checked, if in the sliding window of the width WINDOW_WIDTH, the number of values < WALK_THRESHOLD was at most WINDOW_TOLERANCE.
     - Once there were > WINDOW_TOLERANCE such values, this was the last window.
     - The end of the segment was the last value ≥ WALK_THRESHOLD in the last window.
   - Segments with short duration were discarded.
     - If the number of minute values ≥ WALK_THRESHOLD in the segment was < MIN_DURATION, then the segment was ignored.
2. Second, within the identified segments, subsegments that were suspected to be running bouts were identified.
   - Similar window-based logic was used with subsegments instead of segments and with RUN_THRESHOLD instead of WALK_THRESHOLD.
   - If the number of minute values ≥ RUN_THRESHOLD in the subsegment was ≥ MIN_DURATION, then the subsegment was identified as a running bout and, in the table, was set.
     - column "activity" = "run",
     - column "start" = start of running bout,
     - column "end" = end of running bout.
3. Third, the intervals were classified within the identified segments outside the running bouts. These intervals were walking bouts, or they were too short.
   - The interval of length ≤ WINDOW_TOLERANCE might occur only at the edges of the segment (otherwise, it would be classified as running bout in the second phase.) Such an interval was too short.

- The interval of length from WINDOW_TOLERANCE+1 up to MINDOBA-1 was no longer classified as a running bout, and at the same time, it was too short to be classified as a walking bout.
- If the number of minute values ≥ RUN_THRESHOLD in the interval was < MIN_DURATION, then the interval was ignored.
- Otherwise, the interval was identified as a walking bout, and in the table was set as such:
  - column "activity" = "walk",
  - column "start" = start of walking bout,
  - column "end" = end of walking bout.

Activity conflict resolution

Detected activity segments were cross-referenced with the file DAILY_ACTIVITY_ SUMMARY_ACTIVITY.csv. If, in the file, there was an activity overlapping with our detected activity, then there were two possibilities.

- If the overlapping activity had a name beginning with "Run", "Wal", "Tre", or "Chů" (this corresponds to the activity of running, walking or treadmill), then the detected activity was retained.
- Otherwise, the detected activity was ignored.

Spatiotemporal Parameter Calculation

Each valid activity bout was characterized by:

- **duration** = end – start + 1 (min) = number of minutes in the bout
- **minutes_R** = Number of minutes in which steps/min ≥ RUN_THRESHOLD, i.e. number of minutes of running
  - It is possible that some minutes of running could occur during a bout of walking activity, or that some minutes of walking could occur during a bout of running.
  - Based on the detection algorithm, there had to be minutes_R ≥ MINDOBA during the running activity.
- **minutes_W** = Number of minutes in which WALK_THRESHOLD ≤ steps/min < RUN_THRESHOLD, i.e. number of minutes of walking
  - Based on the detection algorithm, during the walking activity, there had to be minutes_R + minutes_W ≥ MIN_DURATION, but also during the walking activity, there might be minutes_W < MIN_DURATION.
- **steps_count_R** = Total steps in minutes_R
  - Based on the detection algorithm, there had to be steps_count_R > 0 during the run activity.
- **steps_count_W** = Total steps in minutes_W
  - Based on the detection algorithm, there had to be steps_count_W > 0 during the walk activity.

Distance Estimation via Regression Models

The distance was estimated by regression-based modeling of step length (d, cm) as a function of step cadence (p, steps/min.) using two types of models:

- GPS-derived regressions (walking WG and running RG) and
- Biomechanical regressions (walking WB and running RB).

GPS-Based Regressions (WG and RG Models)

Based on the GPS data, the dependence of the step length (d) on the number of steps per minute (p) was estimated for running and walking. The problem was that GPS data sometimes contained a large amount of data that had not been measured but had been calculated. For example, there was unnatural amount of data on the curves d = const and dp = const in the data. Also, there were outliers in the data with impossible step lengths. All such data had to be discarded first.

Therefore, the data were filtered according to the following rules:

- Only data with 20 ≤ step lengths ≤ 200 cm were retained.
- Only data with WALK_THRESHOLD ≤ step cadence < RUN_THRESHOLD were retained for walking bouts.
- Only data with RUN_THRESHOLD ≤ step cadence ≤ MAX_STEPS were retained for running bouts.
- Outlier curves (d=const, dp=const) were detected based on the following:
  - There were at least two times more values on this curve than the number of values on the fourth most dense curve of the same type,
  - There were at least 10 values on this curve.
- All data on the outlier curves were discarded.
- Each particular GPS model was calculated if there were at least 10 non-discarded input values.

For each activity type, models were fitted separately:

- **d = WG0(p)** = a_0 (constant regression for walking), constant step length
- **d = WG1(p)** = $a_1 + b_1 \cdot p$ (linear regression for walking)
- **d = RG0(p)** = a_2 (constant regression for running), constant step length
- **d = RG1(p)** = $a_3 + b_3 \cdot p$ (linear regression for running)

Biomechanical-Based Regressions (WB and RB Models)

Based on biomechanical data, the dependence of the step length (d) on the number of step cadences (p) was estimated for running and walking.

For each activity type, models were fitted separately:

- **d = WB0(p)** = a_4 (constant regression for walking), constant step length
- **d = WB1(p)** = $a_5 + b_5 \cdot p$ (linear regression for walking)

info@osu.eu / +420 553 461 001 / www.osu.eu
University of Ostrava / Dvořákova 7 / 701 03 Ostrava / Czech Republic

- **d = RB0(p)** = a_6 (constant regression for running), constant step length
- **d = RB1(p)** = d = $a_7 + b_7 \cdot p$ (linear regression for running)

If there were no biomechanical data for some regression model (<1 for constant regression, <2 for linear regression), then this model was omitted for that participant.

Merging GPS and BM Models

While the biomechanical regression models were available for a vast majority of participants, the GPS regression models were available only for a portion of them (WG model: 813 (62%) participants; RG model: 671 (51%) participants). On the other hand, the GPS regression models are expected to be more accurate. Therefore, the strategy was to compare both models for participants with both models available, find a relation between both models, and for all participants use the BM model adjusted by that relation.

For each participant with all the walking regression models available, the estimated walking distance was computed for every week using each of these models. These estimates were compared with each other. Similarly for running distances.

In Figure 1, each dot represents one participant and week. On the x-axis, there is an estimated running distance using the RB0 model. On the y-axis, there is an estimated running distance using the RG1 model divided by an estimated running distance using the RB0 model. The value 0.888 is the mean value of the quotient RG1/RB0.

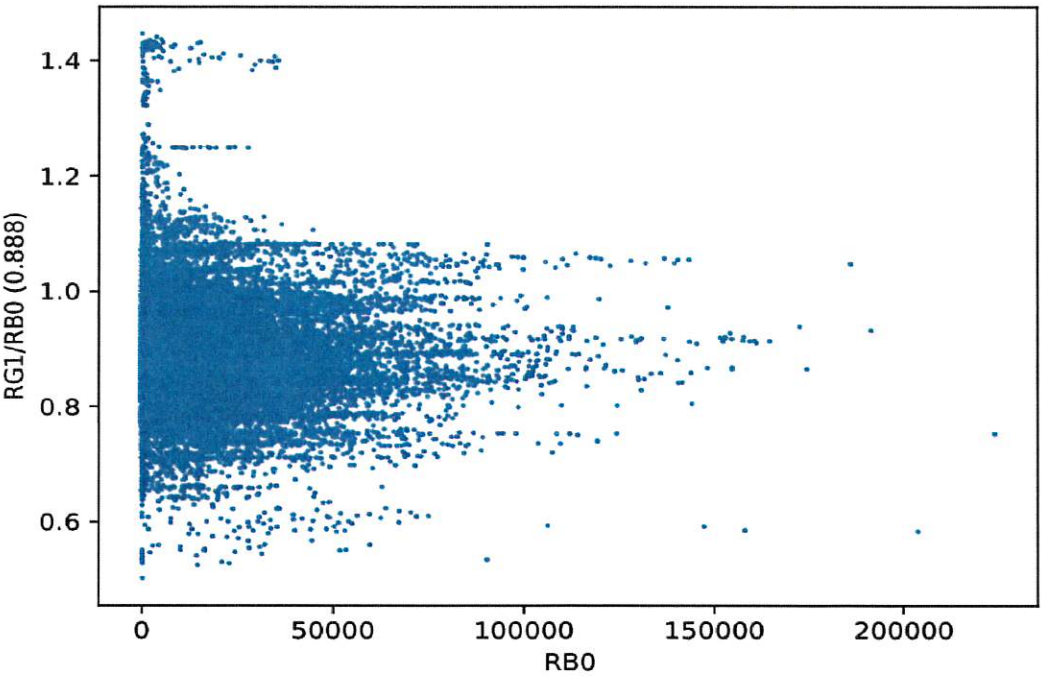


Figure 1 Comparison of RB0 and RB1 models for running distance estimation

The tables (Table 1,2) show mean values of the quotients WGi/WBj and RGi/RBj. Notice that the values practically do not depend on i.

t	WG0/t	WG1/t
WB0	0.928	0.922
WB1	0.977	0.970

Table 1 Mean values of WGi/WBj

T	WG0/t	WG1/t
WB0	0.928	0.922
WB1	0.977	0.970

Table 2 Mean values of RGi/RBj

It does not matter whether WB0 or WB1 for week aggregations were used because the week-estimated walking distance will only be adjusted by a different factor to the same value. On the other hand, for shorter time intervals, it is better to use the WB0 model because it oscillates less than the WB1 model, similarly for running.

Moreover, with respect to the noise in the GPS data, two decimal digits in the quotients are quite sufficient, and rounding to two decimal digits will not decrease the accuracy of the models.

All these points were used to compute values "distance_W" and "distance_R" as described in the subsequent paragraphs.

Column "distance_R"

The distance_R column stores an estimate of the total number of meters during those minutes of running activity that were counted in minutes_R. The number of meters was defined as the sum of minute distances v = 0.89 x p xRB0(p) over all minutes in this activity that were counted in minutes_R. If no running biomechanical data was available for the participant, distance_R was set to 0. The algorithm implied that during the running activity, one of the following events had to occur:

- steps_count_R > 0,
- There was no biomechanical data available for running.

Column "distance_W"

The distance_W column stores an estimate of the total number of meters during those minutes of walking activity that were counted in minutes_W. The number of meters was defined as the sum of minute distances v = 0.93 x p x WB0(p) across all minutes in this activity that were counted in minutes_R. If no running biomechanical data was available for the participant, distance_W was set to 0. The algorithm implied that during the running activity, one of the following events had to occur:

- steps_count_W > 0,
- There was no biomechanical data available for walking.

Speed Estimation

For each activity bout, average speed of running (km/h) and average speed of walking (km/h) were computed.

Column "avg_speed_R"

The avg_speed_R column stores the average speed during those minutes of the activity bout counted in minutes_R, i.e., the value 0.06x(distance_R/minutes_R). The coefficient 0.06 converts metres per minute to kilometres per hour. If minutes_R = 0 then was set avg_speed_R = 0.

Column "avg_speed_W"

The avg_speed_W column stores the average speed during those minutes of the activity bout counted in minutes_W, i.e. the value 0.06x(distance_W/minutes_W). The coefficient 0.06 converts metres per minute to kilometres per hour. If minutes_W = 0 then was set avg_speed_W = 0.

Cardiovascular Metrics

Heart rate zones were calculated by normalizing instantaneous heart rate (HR) relative to each participant's resting HR and maximum HR (from the graded exercise test).

The resting heart rate was in the file RESTING.csv for every day. If for some day the participant's resting heart rate value was missing, then the value from the nearest previous day was taken.

Each HR reading was normalized:

- **q = (HR_actual – HR_rest) / (HR_max – HR_rest)**

The "time_HRzone_*" columns show the number of seconds during the activity that the participant spent in each heart rate zone. All of the seconds from "duration" minutes were counted, i.e., also from minutes that were not counted in minutes_W or minutes_R.

- For q < 0.3, the seconds were counted in time_HRzone_very_light.
- For 0.3 ≤ q < 0.4, the seconds were counted in time_HRzone_light.
- For 0.4 ≤ q < 0.6, the seconds were counted in time_HRzone_moderate.
- For 0.6 ≤ q < 0.9, the seconds were counted in time_HRzone_vigorous.
- For q ≥ 0.9, the seconds were counted in time_HRzone_maximal.

If the value of q was unknown, the seconds were counted in time_HRzone_unknown. This could occur

- in minutes where there was no heart rate record,
- in the days before the first day with the specified resting_HR.

Elevation gain estimation

The "elevation" column shows the elevation meters that the participant climbed during the whole activity. The elevation from all duration minutes was counted, i.e., from minutes that were not counted in minutes_W or minutes_R. For each participant, the input information about the elevation was given in the ELEVATION.csv file. In this file, the elevation values were measured with an accuracy of 10 feet, and this was rounded to millimeters. Thus, the smallest positive value in the ELEVATION.csv file is 3.048 meters. In the "elevation" column, the values are rounded to meters. Therefore, rounding errors can cause a situation where the following two values are not equal:

- the number of meters of participants' elevation in the weekly aggregation,
- the sum of the number of meters of participants' elevation overall activities in that week.

Device Wear-Time Estimation

Wear time was computed at daily, weekly, and yearly aggregation levels according to the following:

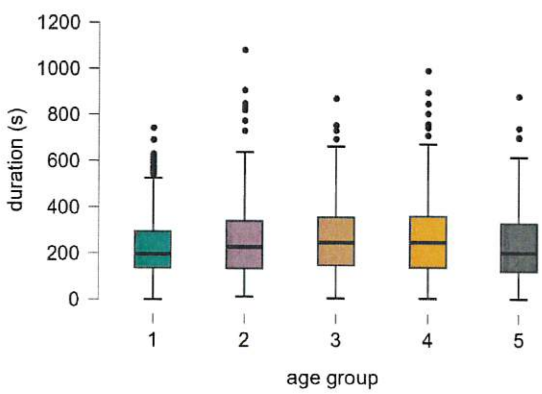
- wear_minutes: Total number of minutes with either step data or HR data, i.e., minutes with
  - at least one recorded step according to the STEPS.csv file or
  - at least one heart rate record according to the HR.csv file.
- wear_days: Number of days per week with wear_minutes > 0.

Results

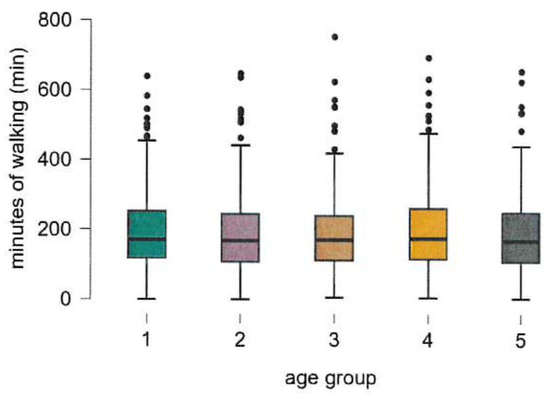
The total sample of 1310 participants was used for the final analysis using the above-mentioned methodological approaches. Four participants from the original sample were excluded due to having no Fitbit data. The mean week values for each participant are presented in figures (Figures 3, 4) displaying means/medians with standard deviations/interquartile ratios across all computed parameters per age group and activity status (runner/non-runner).

Figures 3 – Age groups

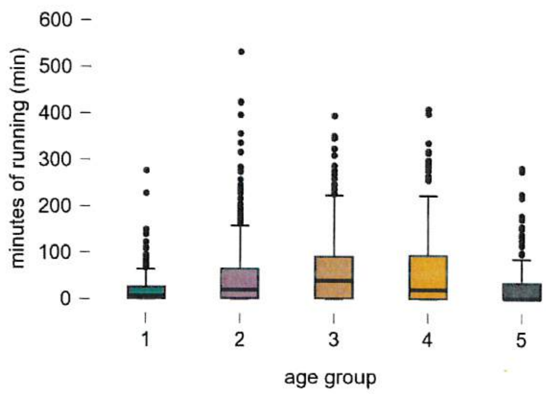
Duration (s)



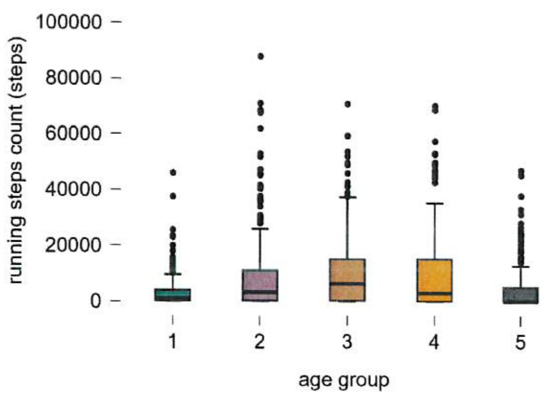
Minutes of walking (min)



Minutes of running (min)



Running steps count (steps)

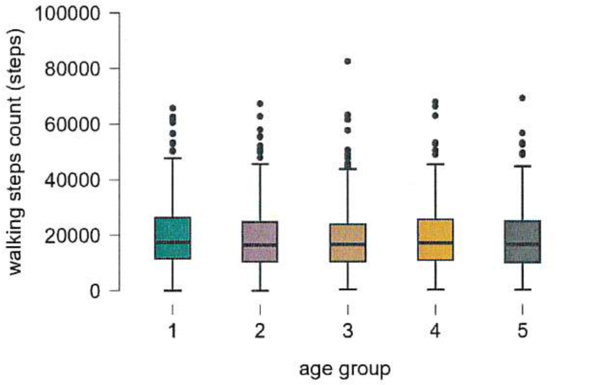


Co-funded by the European Union

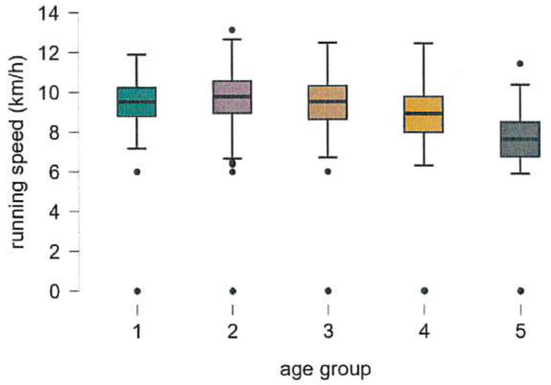
Ministry of the Environment of the Czech Republic

LERCO

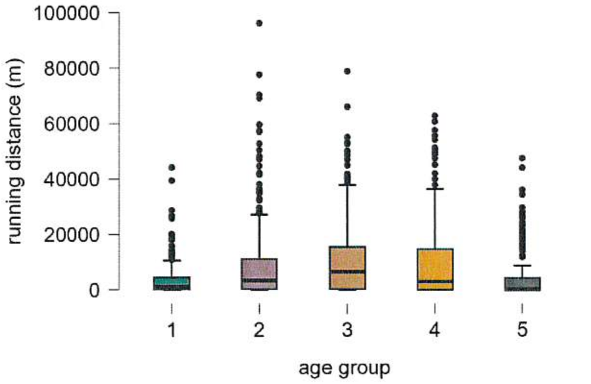
Walking steps count (steps)



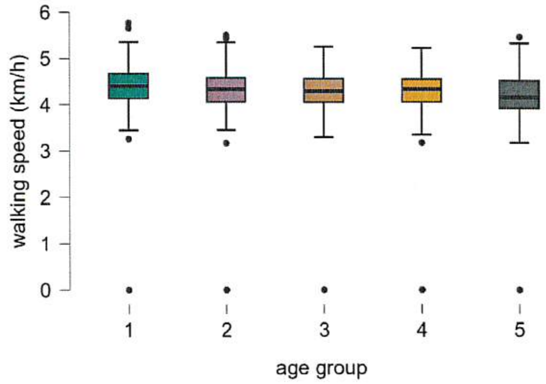
Running speed (km/h)



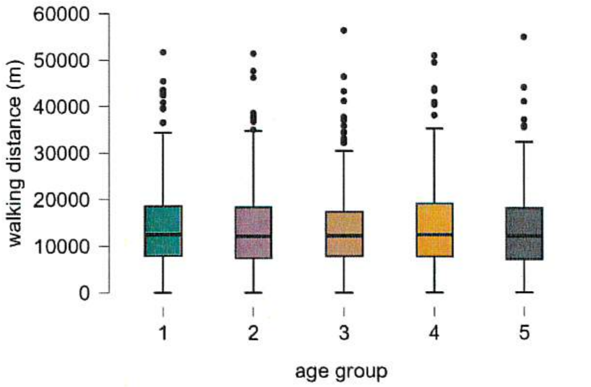
Running distance (m)



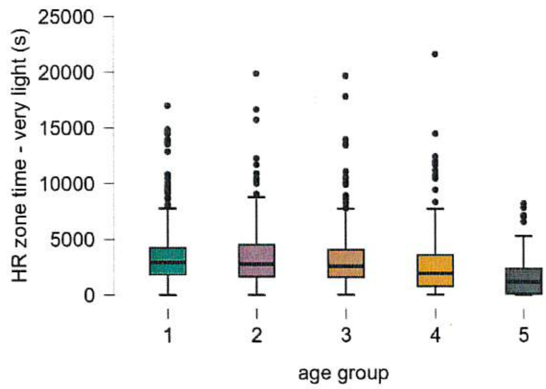
Walking speed (km/h)



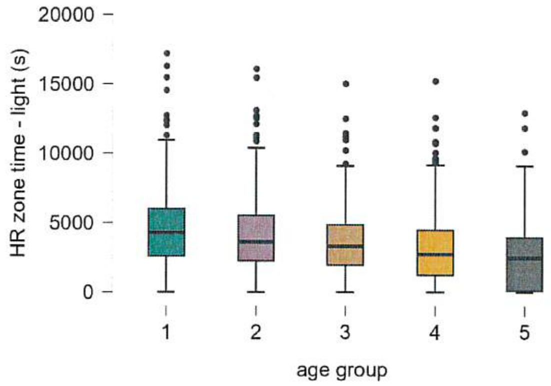
Walking distance (m)



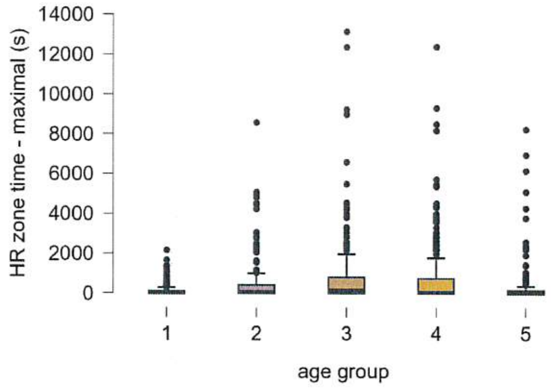
HR zone time - very light (s)



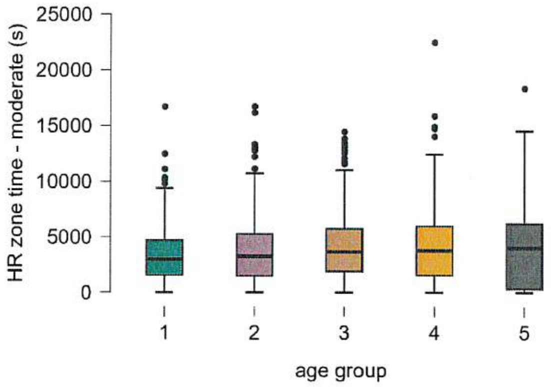
HR zone time - light (s)



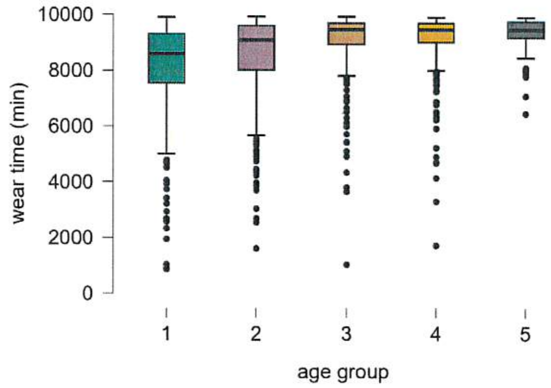
HR zone time - maximal (s)



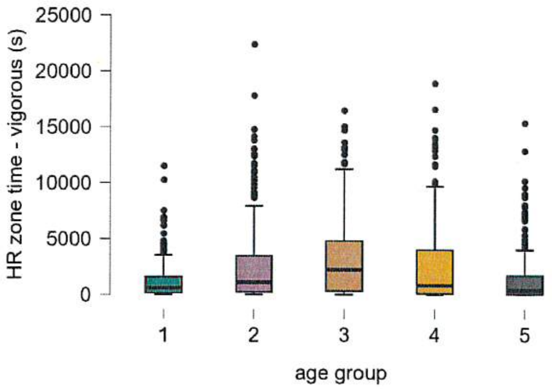
HR zone time - moderate (s)



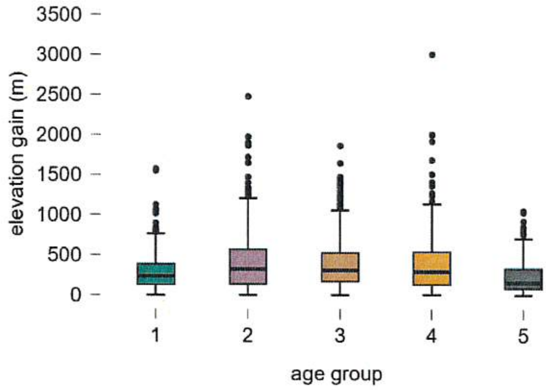
Wear time (min)



HR zone time - vigorous (s)

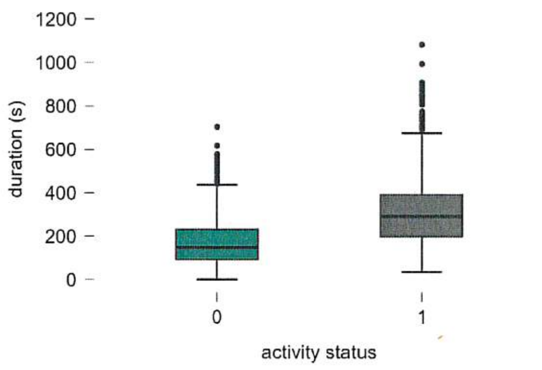


Elevation gain (m)

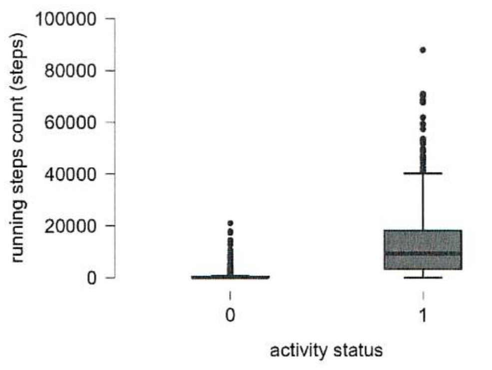


Figures 4 - Activity status

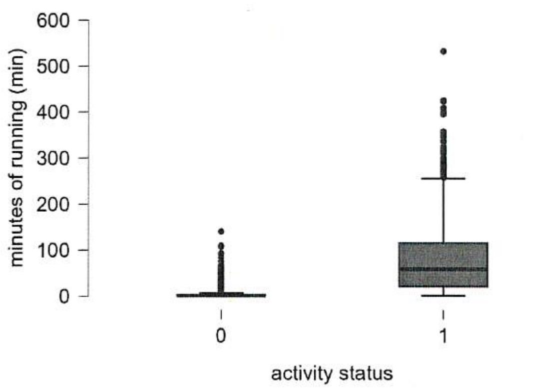
Duration (s)



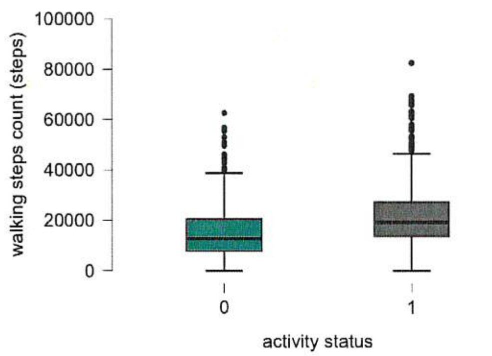
Running steps count (steps)



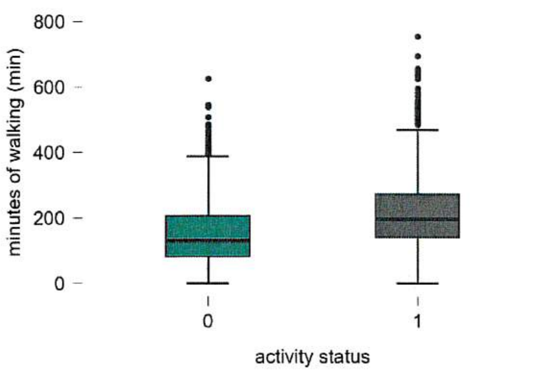
Minutes of running (min)



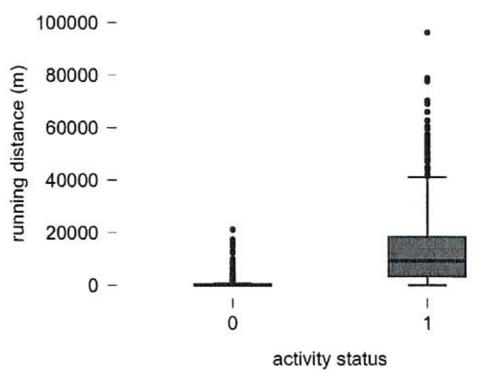
Walking steps count (steps)



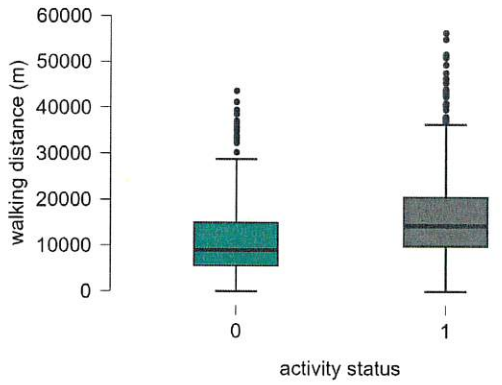
Minutes of walking (min)



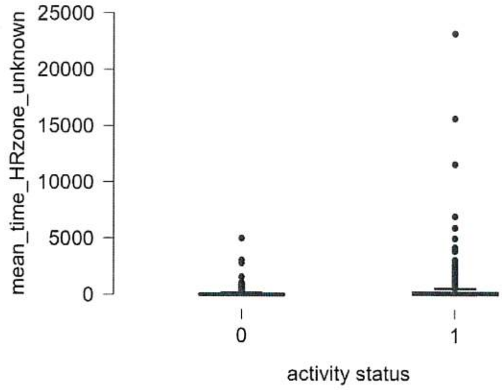
Running distance (m)



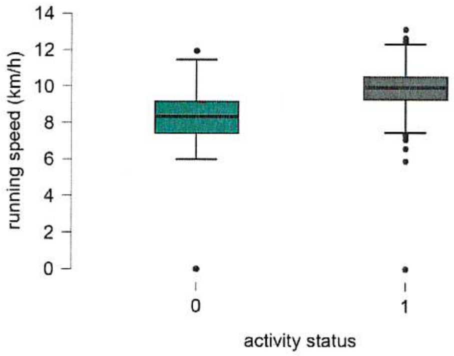
Walking distance (m)



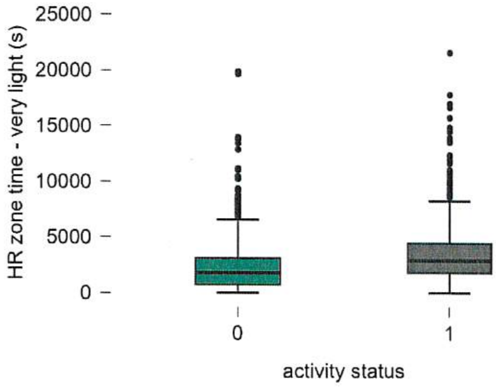
Mean_time_HRzone_unknown



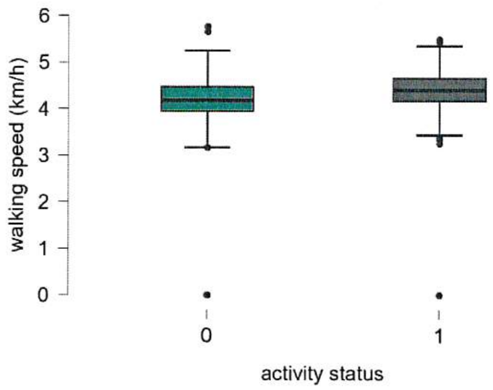
Running speed (km/h)



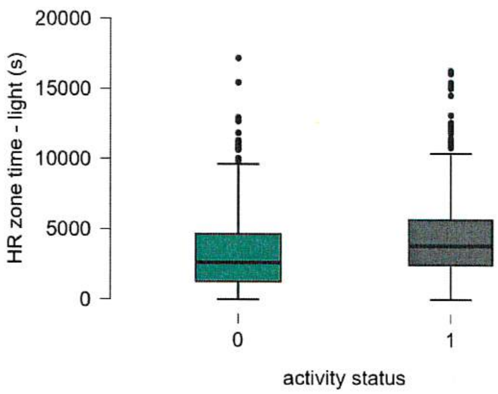
HR zone time - very light (s)



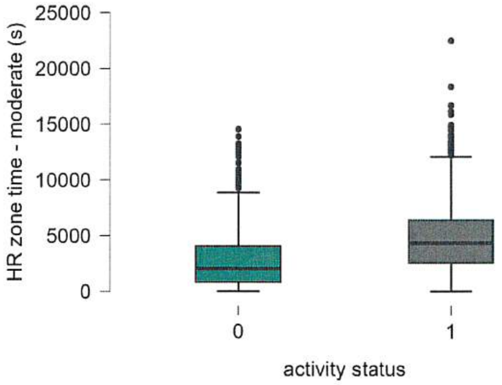
Walking speed (km/h)



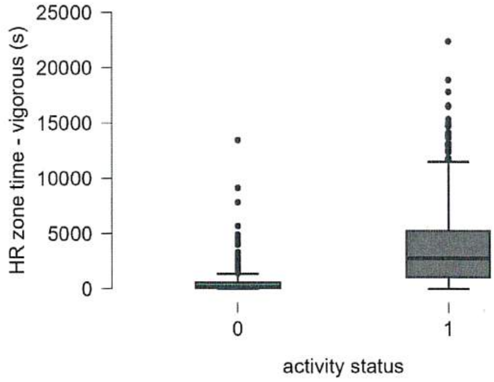
HR zone time - light (s)



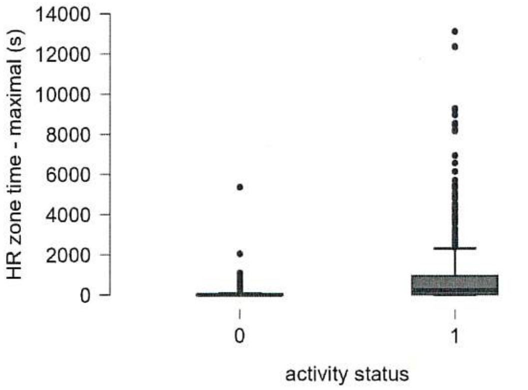
HR zone time - moderate (s)



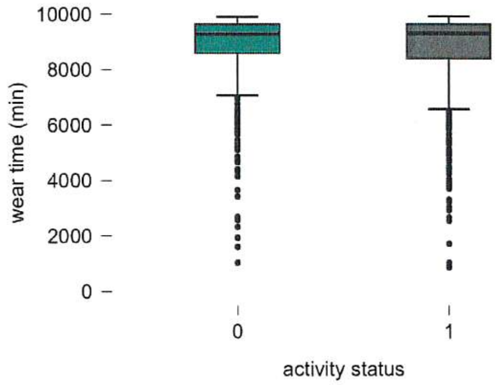
HR zone time - vigorous (s)



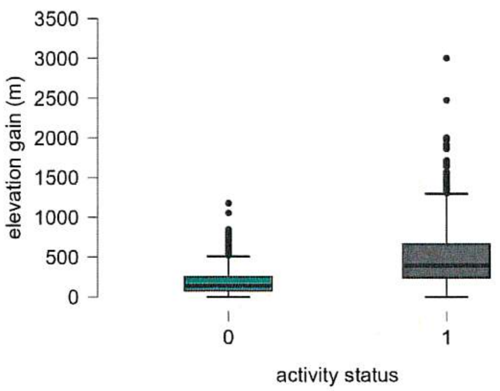
HR zone time - maximal (s)



Wear time (min)



Elevation gain (m)



Conclusions

In this methodological study, we used advanced statistical methods to integrate data from consumer-grade wearable technology with biomechanical laboratory data. Our objective was to estimate the spatiotemporal parameters from 1310 participants of the 4HAIE study who were wearing Fitbit wristbands during one-year follow-up after baseline measurement. We found a strong correlation between the computed walking and running data and the available GPS data. Consequently, we propose that these data can be utilized to estimate the day-to-day, weekly, and yearly amounts of walking and running-related parameters presented in this study

Acknowledgments

Authors' Contribution statement:

Jan Šustek: Software (lead); methodology (lead); formal analysis (lead); writing – review and editing (equal). **Jan Urbaczka:** Conceptualization (equal); software (supporting); methodology (supporting); formal analysis (supporting); writing – original draft (lead); writing – review and editing (equal). **Daniel Jandačka**: Conceptualization (equal); writing – review and editing (supporting). **Steriani Elavsky**: Methodology (supporting); writing – review and editing (supporting).

This methodological approach has been produced with the financial support of the European Union under the project LERCO (CZ.10.03.01/00/22_003/0000003) via the Operational Programme Just Transition as well as under the project „Research of Excellence on Digital Technologies and Wellbeing"(CZ.02.01.01/00/22_008/0004583).

Approved

Approved: prof. Mgr. Daniel Jandačka, Ph.D., leader RP4