



## OPEN Validation and comparative study of the Motus system for accurately identifying movement behaviours using different sampling frequencies

Tonje Pedersen Ludvigsen<sup>1</sup>✉, Sebastian Asmussen Sode Hørlück<sup>1</sup>, Jon Roslyng Larsen<sup>1</sup>, Christina Bach Lund<sup>2</sup>, Pasan Hettiarachchi<sup>3</sup>, Mikkel Brandt<sup>1</sup> & Nidhi Gupta<sup>1</sup>

Accurate data on movement behaviours is essential for public health policy and interventions. Accelerometer-based tools like ActiPASS, paired with AxivityX3 accelerometers, are widely used but face challenges like slow data processing and administrative burden. To address these issues, we developed Motus (alias SurPASS), a scalable accelerometer-based system with wireless SENSmotionPlus accelerometers and cloud storage. While Motus shows promising feasibility, validating its performance at different sampling frequencies can further improve battery life, data transfer, and storage. This study aimed to validate Motus in identifying movement behaviours at different sampling frequencies and to compare it with ActiPASS. Data collection included laboratory and free-living sessions, during which participants wore three thigh-mounted accelerometers: SENSmotionPlus (25 Hz and 12.5 Hz) and AxivityX3 (25 Hz). The laboratory session was video recorded and comprised structured and semi-structured movement behaviours, including sedentary, standing, walking, stair climbing, running, and cycling. In the free-living session, participants wore accelerometers for 48 h and carried out usual activities. Data were processed using Motus, with free-living sessions additionally analysed with ActiPASS. The F1-score and balanced accuracy were used to assess Motus performance against video annotations, while Bland-Altman plots evaluated agreement between F1-scores (25 Hz vs. 12.5 Hz) and between Motus (25 Hz and 12.5 Hz) and ActiPASS, with ActiPASS as the reference method. Eighteen participants (61% female, age  $34.1 \pm 8.3$  years) completed the data collection. A total of 395.7 min from the laboratory (mean:  $22.0 \pm 2.7$  min) and 501,122 min from the free-living session (mean:  $2638 \pm 499.7$  min) were analysed. In the laboratory, overall F1-score and balanced accuracy were 0.94 for both frequencies, compared with video observations. Additionally, mean bias F1-scores were  $\pm 0.01$  for all behaviours when comparing the performance of Motus 12.5 Hz vs. 25 Hz against videos. In free-living, Motus 25 Hz vs. ActiPASS showed a mean difference of  $\pm 1$  min for all behaviours. For Motus 12.5 Hz vs. ActiPASS, the mean difference were  $\pm 1$  min for sedentary, stair climbing and cycling, while a mean difference of +5.1, -2.9 and -2.2 min were observed for standing, walking and running, respectively. In conclusion, our study provides significant insights into the performance of Motus, highlighting the possibility of reducing the sampling frequency from 25 to 12.5 Hz and still obtaining high performance in classifying different movement behaviours.

**Keywords** Accelerometry, Wearable, Surveillance, Thigh-worn, SurPASS

### Abbreviations

ProPASS Prospective physical activity and sitting and sleep consortium  
SD Standard deviation

<sup>1</sup>National Research Centre for the Working Environment, Copenhagen, Denmark. <sup>2</sup>Department of Occupational and Social Medicine, Holbaek Hospital, Holbaek, Denmark. <sup>3</sup>Department of Medical Sciences, Occupational and Environmental Medicine, Uppsala University, Uppsala, Sweden. ✉email: xtpl@nfa.dk

Sed Sedentary  
LoA Limits of agreement

Large-scale and accurate data on movement behaviours, such as physical activity and sedentary behaviours, are essential for informing effective public health interventions, guiding policy development, and informing evidence-based recommendations<sup>1</sup>.

Wearable devices, such as accelerometers placed on the thigh, hip, or wrist, enable continuous movement behaviour monitoring over multiple days, supported by highly accurate processing software<sup>2</sup>. The Prospective Physical Activity, Sitting and Sleep (ProPASS) consortium<sup>3</sup> developed ActiPASS<sup>4</sup>, which integrates the Acti4<sup>5,6</sup> classification algorithm to facilitate offline processing and harmonisation of large-scale accelerometer data across global cohort studies<sup>7</sup>. ActiPASS, the state-of-the-art method, is widely used internationally<sup>8</sup>. However, ActiPASS and similar systems are primarily research-focused, limiting accessibility for non-researchers<sup>4</sup>. Additionally, these systems often experience long data processing delays and impose substantial administrative and participant burdens<sup>9–11</sup>.

In response to these challenges, we have developed the Motus system (alias SurPASS)<sup>12,13</sup> to eliminate in-person meetings, replace paper-based data collection, and automate processing and participant feedback. Motus consists of (1) an easily attachable accelerometer (SENSmotionPlus) with integrated Bluetooth data-transfer technology, (2) a smartphone app that provides participant-instruction and feedback, and transfers data between accelerometer and cloud storage, (3) back-end infrastructure for cloud storage and processing software, (4) processing software, ActiMotus, built upon the validated MATLAB-Acti4 software, and (5) a web-based application for administrators. Recent studies indicate that Motus is highly feasible for large-scale data collection in occupational and public health surveillance<sup>14,15</sup>. However, its reliance on wireless components and cloud storage means that data resolution and volume significantly affect battery life, storage capacity, and transfer time<sup>16</sup>. Investigating the performance of Motus at lower sampling frequencies is therefore crucial to enhancing feasibility, as it could balance performance with practical constraints such as battery life, transfer time, and storage capacity.

Traditionally, systems using accelerometry employ sampling frequencies between 30 and 100 Hz<sup>17</sup>. Recent research shows that lower frequencies<sup>18–20</sup>, below 30 Hz (as used in systems like Motus and ActiPASS<sup>5–7</sup>), are sufficient for monitoring activities of daily living, which rarely exceed 5–6 Hz<sup>21–23</sup>. In fact, Khusainow et al.<sup>20</sup>, using machine learning-based classifiers, found meaningful classification performance of movement behaviours (accuracy of  $\geq 93\%$ ), even at sampling frequencies as low as 10 Hz. Similar accuracy has been reported for no-code classification methods using 12.5 Hz<sup>23–25</sup>.

Although Motus currently uses a sampling frequency of 25 Hz, recent evidence suggests that it may be possible to reduce it further without compromising performance, potentially enhancing the feasibility of Motus. Thus, our study aimed to evaluate the performance of the Motus system at various sampling frequencies.

The specific objectives are to:

- 1) Validate Motus against video observations in identifying movement behaviours at 25 Hz and 12.5 Hz sampling frequencies under laboratory conditions.
- 2) Compare Motus using 25 Hz and 12.5 Hz sampling frequencies against the commonly used ActiPASS tool in identifying movement behaviours in free-living conditions.

## Method

### ActiMotus and ActiPASS and their key similarities and differences

The processing software for Motus, ActiMotus version 1.1.0<sup>26</sup>, and ActiPASS version 2024.05<sup>4</sup> are both built upon the validated MATLAB-based Acti4 processing software<sup>5,6,27</sup>, which classifies thigh-worn accelerometry into movement behaviours. Acti4 employs a decision-tree method based on movement intensity and accelerometer orientation relative to gravity. Raw triaxial accelerometry data undergo 4th-order Butterworth low-pass filtering before movement intensity (mean acceleration, standard deviation) and orientation (forward/backward inclination) are calculated in 2-second windows with 1-second overlap. Acti4 then classifies movement behaviours into stationary and active categories. Within the stationary category, it distinguishes between standing and sedentary (including lying and sitting) based on orientation. The active category is further divided into specific activities like walking, running, stair climbing, and cycling, using a combination of movement intensity and accelerometer orientation. Acti4 quantifies the duration spent in each movement behaviour, with classification criteria detailed in previous publications<sup>5,28</sup>. While ActiMotus and ActiPASS derive from Acti4, key differences are outlined in the following section. Both are open-source and publicly available, and collaboration to further develop and apply them in future research is welcomed<sup>4,26</sup>.

#### ActiPASS

ActiPASS<sup>4,29,30</sup> is designed for batch-processing large volumes of thigh-worn accelerometer data from multiple brands<sup>7</sup>, including the commonly used Axivity AX3. To facilitate large-scale data handling, ActiPASS automatically corrects accelerometer placement errors and applies both device- and individual-specific calibrations using free-living data<sup>5,31</sup>. For this study, ActiPASS is regarded as the most up-to-date and precise implementation of the validated Acti4 algorithms.

#### ActiMotus

To enhance usability and automation, the MATLAB-based Acti4 software has been translated into Python and integrated into the Motus system's back-end infrastructure as ActiMotus version 1.1.0<sup>26</sup>. The source code is presented in Additional File 1 and available on GitHub (<https://github.com/acti-motus/acti-motus>). It retains

Acti4's core functionalities but is optimised for the Motus environment. A key distinction lies in data processing: in ActiMotus, raw accelerometry data undergo pre-processing in chunks, where movement intensity and accelerometer orientation features are extracted before classification. This approach optimises computational flow, as classification relies on derived features rather than raw data. In contrast, Acti4 directly accesses raw data during classification.

Furthermore, when the SENSmotionPlus accelerometer was introduced into Acti4, a device-specific adjustment was made to align the standard deviation feature with those of more widely used accelerometers, such as Axivity AX3 and ActiGraph GT3X+, upon which Acti4 was originally based. Because SENSmotionPlus has a smaller dynamic range (see specifications in Additional File 2, Table A1), its standard deviation values during high-intensity movements were slightly reduced, likely due to temporary signal saturation. Since the standard deviation is used to define movement intensity thresholds in the ActiMotus decision tree, a second-order polynomial correction was applied during pre-processing to compensate. The implementation of this correction is available in Additional File 1 and the public GitHub repository (<https://github.com/acti-motus/acti-motus>).

## Recruitment and ethics

Participants were recruited from a mid-sized workplace in Denmark, where most held white-collar jobs. The study aimed to enrol 15–25 participants, a typical sample size in similar validation research<sup>5,6,32,33</sup>. Exclusion criteria included allergies to plaster (to avoid adverse skin reactions), pregnancy, or health conditions that could affect normal movement behaviour, including fever on the day of data collection. These criteria were applied to ensure that the system was evaluated under the population conditions for which it was originally developed and validated<sup>5</sup>. In accordance with the Declaration of Helsinki, participants received written and oral study information before providing written informed consent. Ethical approval was granted by the Committee of Scientific Research Ethics for the Copenhagen Region (j.nr. F-24042892). The study complies with national and international data protection regulations.

## Protocol

### Instrumentation

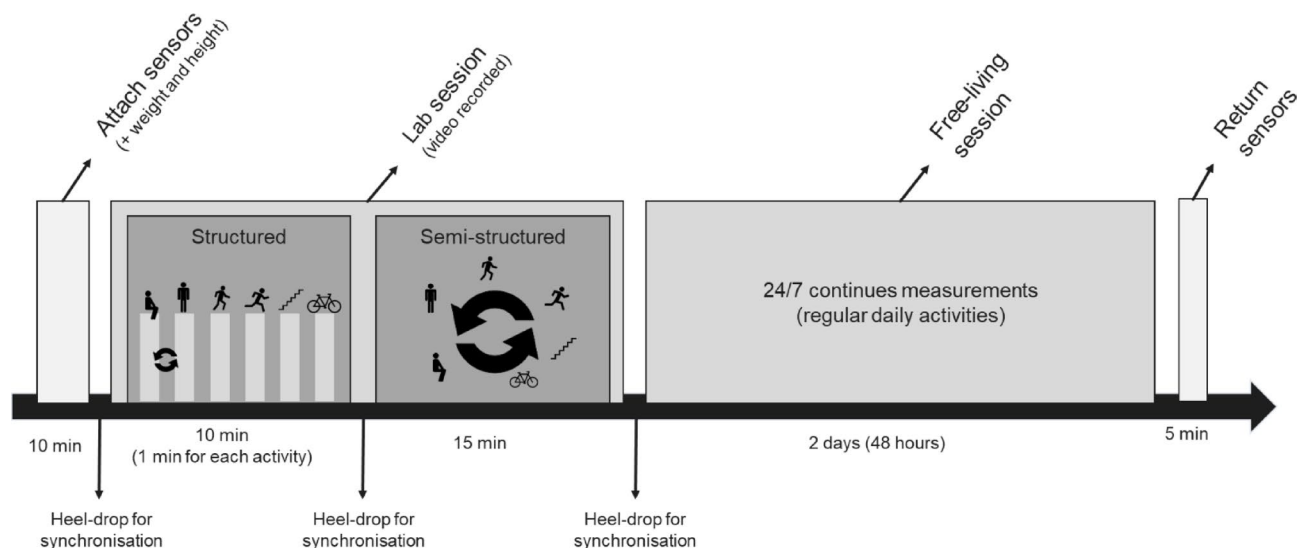
Data were collected using accelerometers from two manufacturers: SENSmotionPlus (SENS Innovation ApS, Copenhagen, Denmark) and Axivity AX3 (Axivity Ltd., Newcastle, UK). Motus employs the wireless SENSmotionPlus, while ActiPASS typically utilises accelerometers like the Axivity AX3. Axivity AX3 data were used solely for free-living comparison as part of ActiPASS (objective 2). Each participant wore two SENSmotionPlus accelerometers, set to 12.5 Hz and 25 Hz, and one Axivity AX3, set to 25 Hz.

### Data collection

Data were collected over three weeks in May 2023. As shown in Fig. 1, the laboratory session consisted of structured and semi-structured components, followed by a free-living session with continuous measurement over two consecutive days.

### Preparation

Before the laboratory session, participants' height and weight were measured. Participants were equipped with three accelerometers placed anteriorly on the midline of the right thigh, specifically at the midpoint between the anterior superior iliac spine and the superior margin of the patella, as visualised in Fig. 2. The



**Fig. 1.** Timeline of data collection, including accelerometer attachment, laboratory and free-living sessions, and device return.



**Fig. 2.** Placement of the accelerometers.

two SENSmotionPlus accelerometers were affixed on top of each other under the same plaster, with the order of placement being randomised. The Axivity AX3 was positioned two centimetres above or below the SENSmotionPlus accelerometers, also in a random order.

#### *Structured and semi-structured session*

In the laboratory, participants completed a session lasting approximately 30 min in indoor and outdoor settings, performing six movement behaviours (sedentary, standing, walking, stair climbing, running, and cycling). They followed a predefined order, beginning with either sedentary (i.e., in the laboratory setting, this means only sitting) or standing to minimise possible order effects and ensure that posture detection was not influenced by always performing one posture before the other. The remaining activities required more setup and were therefore done in a fixed order for practical reasons. Each behaviour lasted one minute, with a 15-second break. The protocol simulated daily activities, e.g., sitting at a computer, walking or running inside and outside. After the structured session, participants engaged in a semi-structured session lasting approximately 15 min, self-selected locations (within the nearby surroundings) and durations for each behaviour. Activities included getting coffee, conversing with colleagues, and cycling in the parking lot. Both sessions were video recorded using a Samsung Galaxy S5 smartphone, held by a research assistant who followed the participant around.

#### *Synchronisation*

Before, during, and after the laboratory session, participants performed three heel-drops while holding the smartphone camera. These heel-drops produced distinct spikes in the accelerometer signal and corresponding shakes in the video, enabling alignment and synchronisation of accelerometer data across devices.

#### *Free-living*

After the laboratory session, participants continued wearing all three accelerometers for 48 h during their normal daily routines. Devices were returned after two days. These data were used for a descriptive comparison of the classified time between recordings at different sampling frequencies (12.5 Hz vs. 25 Hz), with ActiPASS used as a reference method (objective 2). This comparison was not intended as a validation but to explore whether and how sampling frequency may lead to differences in classification results under real-world conditions.

### **Data Preparation**

#### *Video annotation*

Videos were prepared for manual annotation using the Anvil software package<sup>34</sup>. Video footage was downloaded from the smartphone and converted to AVI format (640 × 360 pixels, 30 fps). Annotation followed a predefined coding scheme describing the onset and offset of movement behaviours (Additional File 3, Table A2). These annotations served as the ground truth for validating movement classifications by Motus at 25 Hz and 12.5 Hz.

Each video was independently annotated twice by separate raters to assess inter-rater reliability. The inter-rater agreement for classifying movement behaviours was 97%, with Cohen's kappa ( $\kappa = 0.96$ )<sup>35</sup>.

#### *Synchronisation and data processing of accelerometer signal and video*

To reduce variability introduced by sensor hardware and individual posture differences, a two-step calibration procedure was performed prior to classification and comparison with video annotations. This included

device-specific autocalibration<sup>31</sup> and individual-specific calibration<sup>36</sup>. This procedure is integrated into both the ActiPASS and Motus systems to harmonise data across devices and minimise inter-monitor variability. To compare classifications from the three accelerometers to video annotations (ground truth), the real-time clocks of the accelerometers and video cameras were synchronised, including alignment and clock-drift correction. Synchronisation was achieved using predefined heel-drop anchor points and signal cross-correlation-based drift correction<sup>37–39</sup>.

For the validation of Motus using 25 Hz and 12.5 Hz sampling frequencies against videos, the following movement behaviours were included: sedentary (sit), standing, walking, running, stair climbing, and cycling. Behaviours categorised as “other,” “invisible,” or “uncertain” by the annotators during the laboratory session were excluded from the analysis. Transitions between behaviours were not annotated as a separate category but were naturally included within the defined activities.

For comparing Motus (25 Hz and 12.5 Hz) with ActiPASS, the same movement behaviours were included, along with lying and moving. Lie was analysed with sit (sedentary), since distinguishing between sitting and lying with a single sensor was beyond this study’s scope and has been validated elsewhere<sup>29</sup>. Move, which refers to brief, irregular activities such as shifting, turning, or taking a few steps, was treated as a transitional state between standing and walking, relevant in free-living conditions. However, it was excluded from the laboratory validation due to the difficulty of reliably staging and annotating such behaviour, as it lacks clear start and end points. Free-living data were averaged over 24 h per participant.

#### Statistical analysis

Statistical analyses were conducted in Python 3.11 using Pandas 1.5<sup>40</sup> for data wrangling, NumPy 1.24<sup>41</sup> for various numerical computations, SciPy 1.10<sup>42</sup> for signal processing and SciKit Learn 1.3<sup>43</sup> for performance metrics calculation.

**F1-score and balanced accuracy** System performance was evaluated using F1-score and balanced accuracy for each movement behaviour and overall performance, based on classification data pooled across all participants. These metrics were selected for their widespread use in previous studies, ensuring comparability with past and future research on human activity recognition<sup>5,6,23,33</sup>.

The F1-score for each movement behaviour was calculated as the balanced mean of precision and recall (recall is equivalent to sensitivity). Precision is the proportion of true positives out of all predicted positives, and recall is the proportion of true positives out of all actual positives. In simple terms, precision means how often the system is correct when it says a behaviour is happening, while recall means how often the system detects a behaviour when it actually happens.

The F1-score was then calculated as:

$$F1_{score} = 2 \cdot \frac{precision \times recall}{precision + recall}$$

The overall F1-score was calculated as a weighted average of the F1-scores for each movement behaviour. The F1-score for each behaviour was weighted based on how often that behaviour occurred in a larger study with 340 participants and 6.5 years of data collected using the Motus system<sup>14</sup>. This approach ensures that the overall score reflects how well the system performs on behaviours that are more likely to happen in real-world situations.

For balanced accuracy, we calculated the mean of recall and specificity for each movement behaviour. Specificity refers to the proportion of true negatives out of all actual negatives. True negatives were identified by a ‘one-against-all’ approach. In simple terms, specificity is how often the system correctly says a behaviour is not happening when it really is not. The balanced accuracy for each behaviour was calculated as the average of recall and specificity:

$$Balanced\ accuracy = \frac{recall + specificity}{2}$$

The overall balanced accuracy was calculated as the average of the recall values obtained for each behaviour.

**Confusion matrix** To assess misclassification trends and classification accuracy, a confusion matrix was generated for each movement behaviour and sampling frequency (25 Hz and 12.5 Hz), as well as for overall system accuracy. Confusion matrices were based on classification data pooled across all participants. These matrices evaluate both the proportion (%) and absolute count of correctly classified instances, highlighting common misclassifications.

**Bland-Altman plot** The Bland-Altman plot<sup>44</sup> was used to (a) test the agreement between the F1-scores calculated for different sampling frequencies (25 Hz and 12.5 Hz) and to (b) test the agreement between the minutes spent on various movement behaviours classified by Motus vs. ActiPASS. These analyses were conducted at the participant level, using (a) individual F1-scores and (b) total minutes spent in each movement behaviour. For the agreement between F1-scores, 25 Hz was used as the reference, and for the agreement between Motus and ActiPASS, ActiPASS was used as the reference method and denominator for relative interpretation.

The plot presents the mean bias (mean difference), limits of agreement (mean difference  $\pm$  1.96  $\times$  standard deviation) and general shape of the spread of data. A mean bias close to zero suggests good agreement. A positive or negative bias indicates whether one method tends to overestimate or underestimate the values compared with the other method. Narrow limits of agreement indicate consistent results between sampling frequencies (F1-



score) and systems (minutes), while wider limits suggest greater variability in the results. Regarding the shape of the data spread, a random scatter or no shape indicates consistent differences, while shapes in the scatter indicate proportional biases (e.g., with a higher mean F1-score, the mean difference between two sampling frequencies increases).

Results

Of the 20 participants who consented to participate, 18 (61% female, age  $34.1 \pm 8.3$  years, body mass index  $23.2 \pm 2.5$  kg/m<sup>2</sup>) completed the entire protocol with valid data, resulting in 406.2 min of recorded data from the laboratory session (mean  $\pm$  SD =  $22.6 \pm 2.6$  min) and 50,122 min of recorded data from the free-living session (mean  $\pm$  SD =  $2638 \pm 499.7$  min). Of the two excluded participants, one was removed due to abnormal movement patterns, and the other due to technical issues with the recorded data.

Validation of Motus using different sampling frequencies in a laboratory setting

After excluding data annotated and categorised as “other”, “invisible”, and “uncertain”, from the lab session, the dataset consisted of 395.7 min. Walking was the dominant activity during the laboratory session, with 154.4 min, while running was the least dominant activity with 29.8 min. The total time spent in sedentary behaviour was 46.7 min, standing 94.4 min, stair climbing 34.2 min, and cycling 36.1 min.

Performance metrics

Table 1 shows the performance of Motus using 25 Hz and 12.5 Hz (both SENSmotionPlus) sampling frequencies compared with video observations, in classifying the six targeted movement behaviours (sedentary, standing, walking, running, stair climbing, and cycling). The results demonstrate an F1-score and balanced accuracy of 0.94 for Motus sampling at 25 Hz and 12.5 Hz. Specifically, for Motus sampling at 25 Hz, the F1-score varied from 0.87 (stair climbing) to 0.96 (cycling), and the balanced accuracy varied from 0.95 (sedentary and stair climbing) to 0.99 (running). For Motus sampling at 12.5 Hz, F1-score varied from 0.86 (stair climbing) to 0.96 (standing, walking, running and cycling), and balanced accuracy varied from 0.95 (sedentary and stair climbing) to 0.98 (standing and walking).

Confusion matrix

Figure 3 shows the confusion matrices, with the diagonal line from the top left to bottom right representing the percentage of correctly classified movement behaviours compared with video annotations for both 25 Hz (right) and 12.5 Hz (left). The overall accuracy was 94% and 95% for 25 Hz and 12.5 Hz, respectively. For specific behaviours, sedentary (~89%) and stair climbing (~91%) showed the lowest classification accuracy at both frequencies, with most of the sedentary behaviour misclassified as standing (~9%) or cycling (~2%) and most of the stair climbing behaviour misclassified as walking (~9%). For standing, walking and cycling, similar classification accuracy was seen for both sampling frequencies, with an accuracy between ~94–97%. The classification accuracy for running was close to 100% for 25 Hz, while for 12.5 Hz, running showed a classification accuracy of ~96% with most of the misclassification categorised as walking (~3%) or stair climbing (~1%).

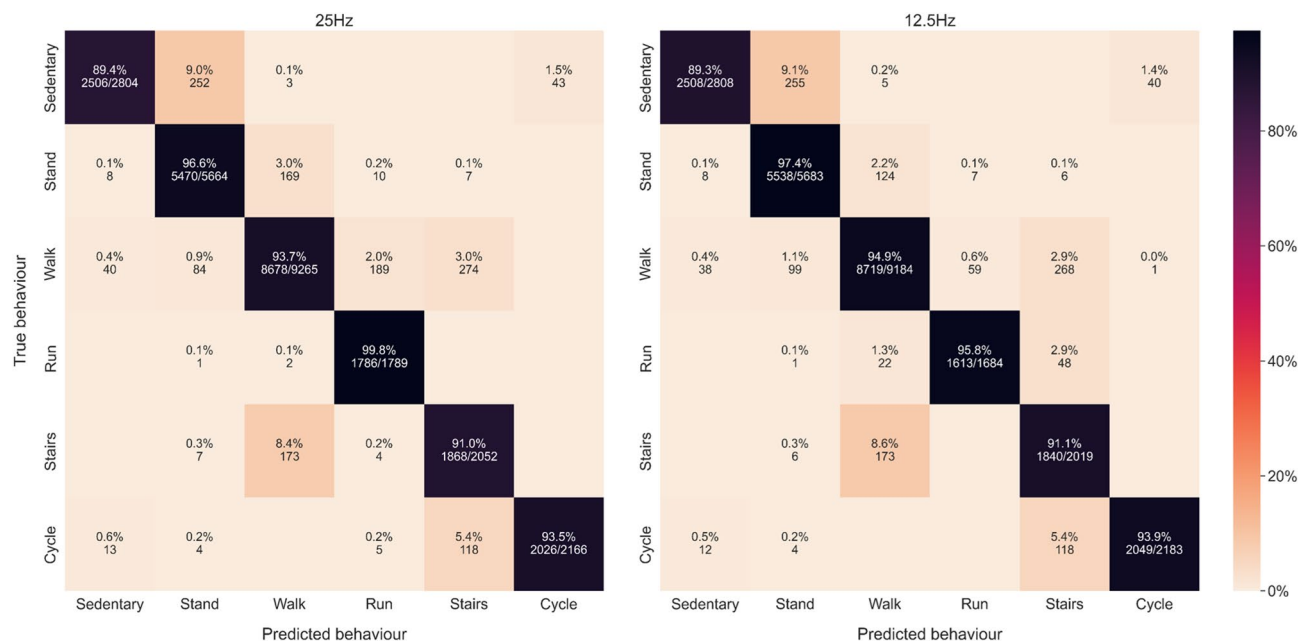
The confusion matrix compares the system’s classification of six behaviours (sedentary, stand, walk, run, stairs, and cycle) against video observations, based on the included sample (n = 18). The Y-axis (rows) represents the true behaviours (based on video annotations), while the X-axis (columns) represents the predicted behaviour (from accelerometer data). Diagonal elements indicate the percentage of correctly classified instances, whereas off-diagonal elements represent misclassifications, where one behaviour was incorrectly classified as another.

Bland-Altman plot of agreement between Motus 25 Hz and Motus 12.5 Hz

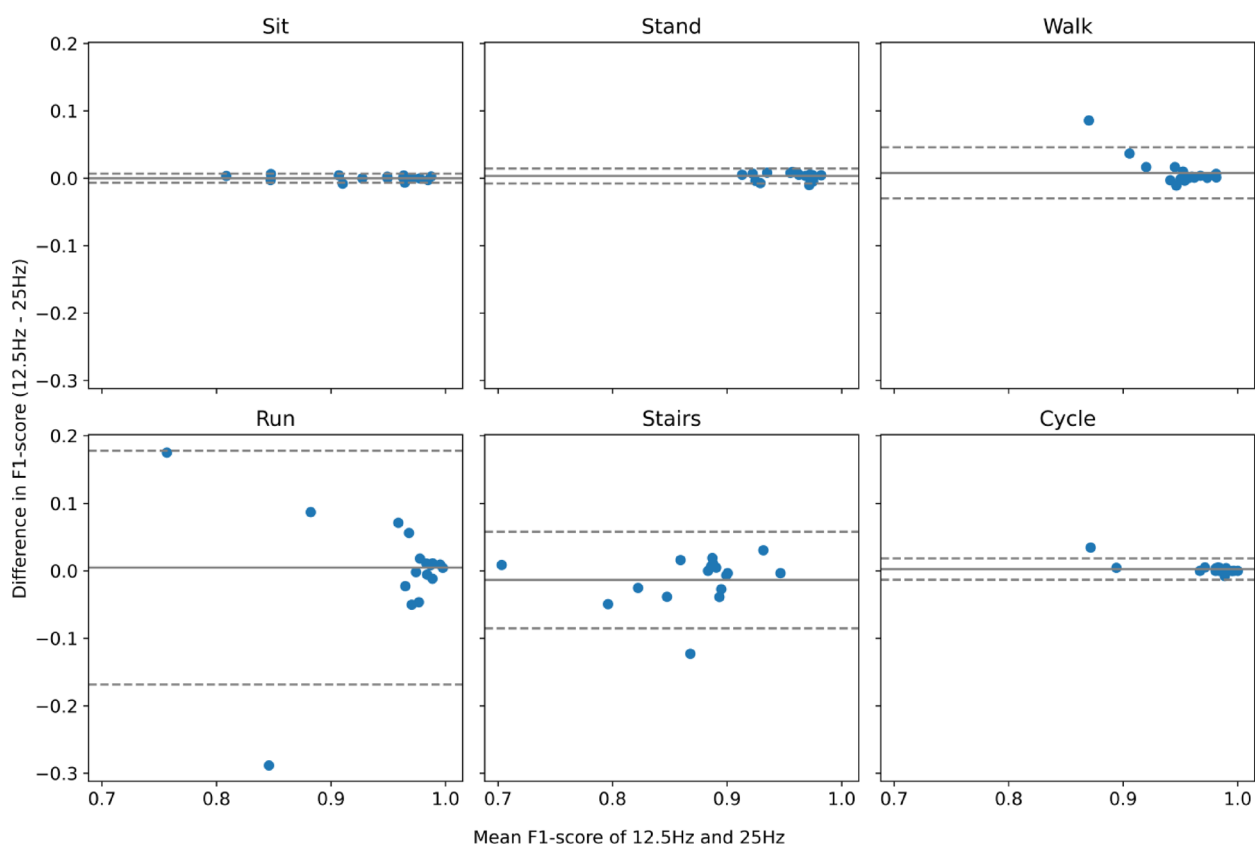
As shown in the Bland-Altman plot, Fig. 4, the mean bias in the F1-score when comparing 12.5 Hz and 25 Hz was close to zero for all behaviours, remaining within  $\pm 0.01$  on a scale from 0 to 1. Of the six behaviours analysed, only stair climbing showed a negative mean bias, indicating a lower F1-score at 12.5 Hz compared with 25 Hz. The limits of agreement (dotted lines) were narrowest for sedentary and standing ( $-0.01$  to  $0.01$ ),

Sampling frequency	Performance metric	Movement behaviour						
		Sed	Stand	Walk	Run	Stairs	Cycle	Overall
25 Hz	Precision	0.98	0.94	0.96	0.90	0.82	0.98	0.96
	Recall	0.89	0.97	0.94	1.00	0.91	0.94	0.93
	Specificity	1.00	0.98	0.98	0.99	0.98	1.00	0.99
	F1-score	0.93	0.95	0.95	0.94	0.87	0.96	0.94
	Balanced accuracy	0.95	0.97	0.96	0.99	0.95	0.97	0.94
12.5 Hz	Precision	0.98	0.94	0.96	0.96	0.81	0.98	0.96
	Recall	0.89	0.97	0.95	0.96	0.91	0.94	0.92
	Specificity	1.00	0.98	0.98	1.00	0.98	1.00	0.99
	F1-score	0.93	0.96	0.96	0.96	0.86	0.96	0.94
	Balanced accuracy	0.95	0.98	0.96	0.98	0.95	0.97	0.94

**Table 1.** Performance metrics for movement behaviour classification by Motus at 25 Hz and 12.5 Hz. Sed = Sedentary (sitting).



**Fig. 3.** Confusion matrix for Motus using SENSmotionPlus at 25 Hz (left) and 12.5 Hz (right).

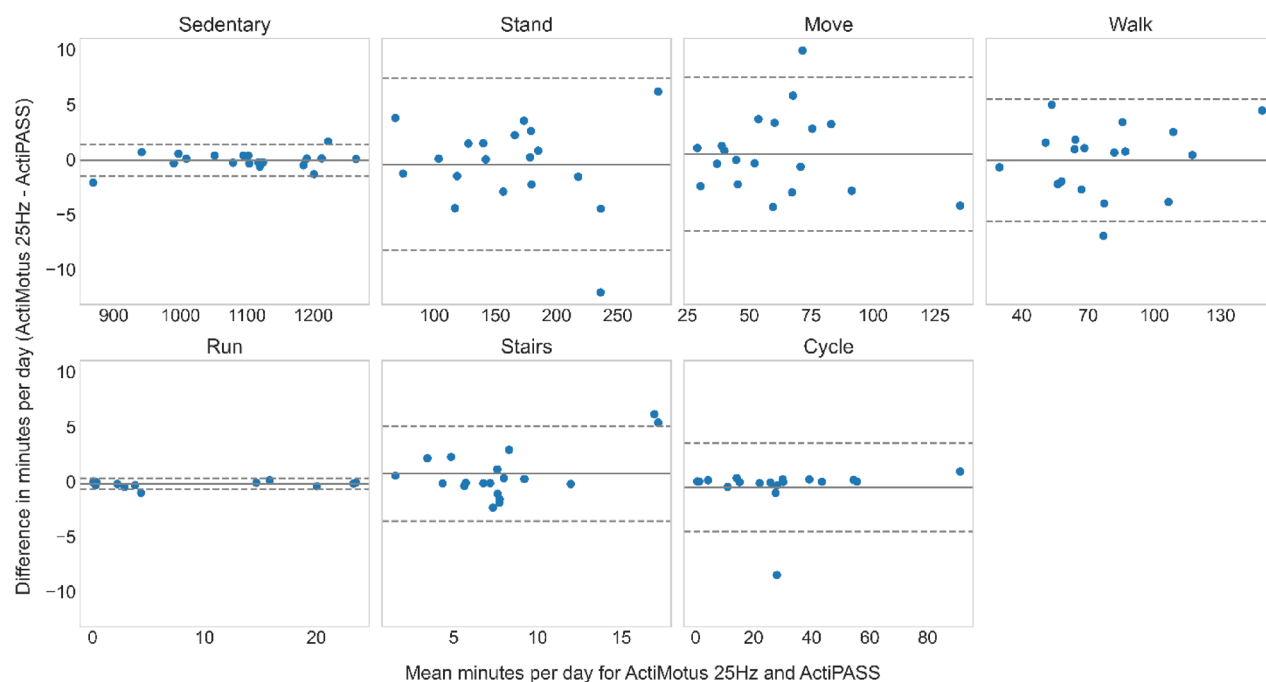


**Fig. 4.** Agreement in F1-Scores for classifying movement behaviours using Motus 25 Hz and 12.5 Hz.

and widest for stair climbing (−0.09 to 0.06) and running (−0.17 to 0.18). For walking, the limits of agreement were −0.03 to 0.05 and for cycling, −0.01 to 0.02. No specific shape of the data spread was seen for any of the movement behaviours.

Movement behaviour	Motus 25 Hz		Motus 12.5 Hz		ActiPASS	
	Mean ( $\pm$ SD)	%	Mean ( $\pm$ SD)	%	Mean ( $\pm$ SD)	%
Sedentary	1098.5 ( $\pm$ 103.8)	76.3	1098.5 ( $\pm$ 103.9)	76.3	1098.6 ( $\pm$ 103.6)	75.7
Stand	162.0 ( $\pm$ 55.7)	11.2	167.5 ( $\pm$ 57.2)	11.6	162.4 ( $\pm$ 56.2)	11.2
Move	61.0 ( $\pm$ 25.2)	4.2	60.8 ( $\pm$ 25.4)	4.2	60.4 ( $\pm$ 25.5)	4.2
Walk	76.6 ( $\pm$ 28.3)	5.3	74.1 ( $\pm$ 27.5)	5.2	76.7 ( $\pm$ 27.7)	5.3
Run	6.1 ( $\pm$ 8.8)	0.4	3.9 ( $\pm$ 7.4)	0.3	6.3 ( $\pm$ 8.8)	0.4
Stairs	8.2 ( $\pm$ 4.7)	0.6	7.6 ( $\pm$ 4.3)	0.5	7.5 ( $\pm$ 3.5)	0.5
Cycle	28.7 ( $\pm$ 22.5)	2.0	28.7 ( $\pm$ 22.5)	2.0	29.2 ( $\pm$ 22.3)	2.0

**Table 2.** Measured minutes of movement behaviours (Mean  $\pm$  SD) over 24 h using Motus 25 Hz, Motus 12.5 Hz and ActiPASS. %Represents the mean percentage of each movement behaviour over 24 h, based on the mean value for each participant. SD = standard deviation.



**Fig. 5.** Agreement in movement behaviour classification (minutes) between Motus 25 Hz and ActiPASS.

The Bland-Altman plot compares the agreement of F1-scores (scale 0–1) between Motus at two sampling frequencies (25 Hz and 12.5 Hz). The mean bias (mean difference) is presented, where a bias close to zero indicates good agreement. A positive bias indicates better performance at 12.5 Hz, whereas a negative bias suggests poorer performance. The dashed lines represent the 95% limits of agreement (mean difference  $\pm$  1.96  $\times$  standard deviation), with narrow limits indicating consistent results and wider limits suggesting greater variability. For visualisation, one participant from the cycling behaviour category was removed as an extreme outlier (F1-scores: 0.00 at 25 Hz and 0.03 at 12.5 Hz), potential reasons for this outlier are elaborated in Strengths and limitations.

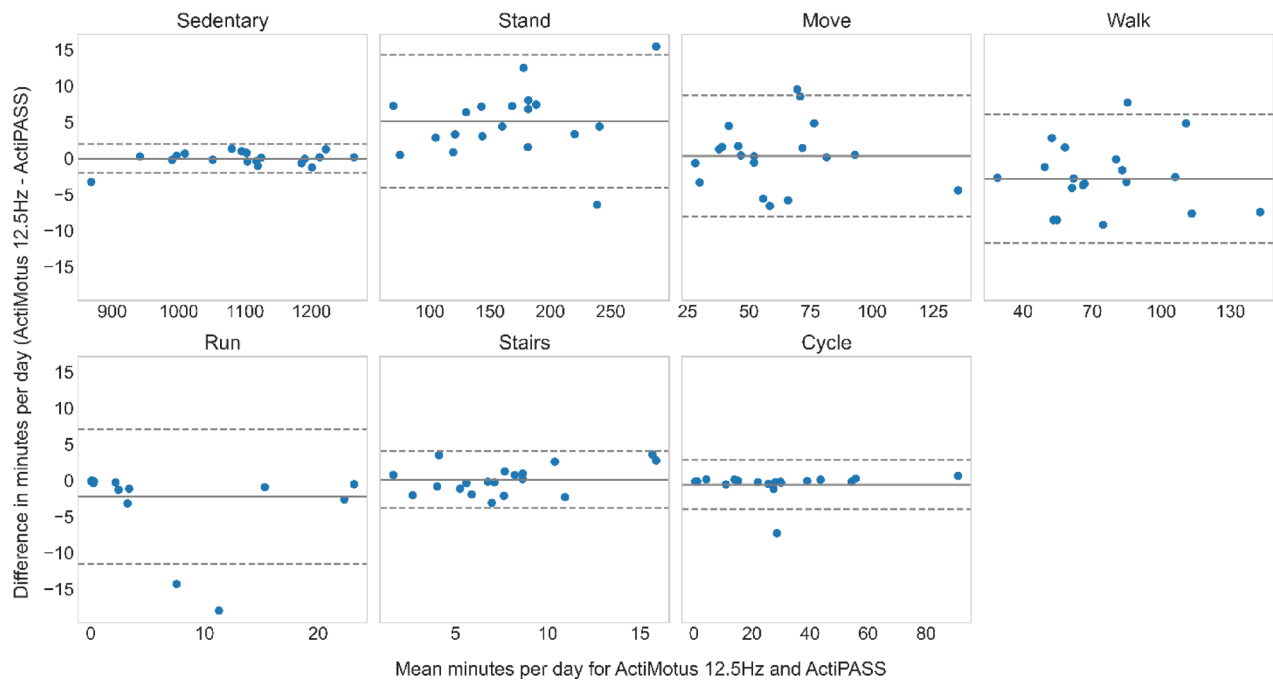
### Comparison of motus and ActiPASS during free-living conditions

Table 2 shows the mean minutes and the percentage of total time for each classified movement behaviour, averaged over 24 h, for both Motus (25 Hz and 12.5 Hz) and ActiPASS. Of the six movement behaviours (sedentary, standing, walking, running, stair climbing, and cycling), sedentary was the most dominant behaviour with over 16 h (1098 min), while running was the least dominant behaviour with less than 7 min per day.

#### Bland-Altman plot of agreement between motus and ActiPASS

For the free-living data, Bland-Altman plots were used to assess the agreement between the classified minutes for Motus (at 25 Hz and 12.5 Hz) and ActiPASS. Figures 5 and 6 compare differences in absolute time spent on each behaviour for each participant, between Motus 25 Hz and ActiPASS, and Motus 12.5 Hz and ActiPASS, respectively. An overview of the mean bias and limits of agreement in minutes, including relative values based on ActiPASS classifications, is presented in Table 3.





**Fig. 6.** Agreement in movement behaviour classification (minutes) between Motus 12.5 Hz and ActiPASS.

Movement behaviour	Motus 25 Hz - ActiPASS			Motus 12.5 Hz - ActiPASS		
	Mean bias (%)	LoA-	LoA+	Mean bias (%)	LoA-	LoA+
Sedentary	−0.1 (−0%)	−1.5	1.4	0.0 (0%)	−2.1	2.0
Stand	−0.4 (−0%)	−8.3	7.4	5.1 (3%)	−4.1	14.3
Move	0.5 (1%)	−6.5	7.5	0.3 (1%)	−8.0	8.7
Walk	−0.1 (−0%)	−5.6	5.5	−2.9 (−4%)	−11.7	6.0
Run	−0.2 (−3%)	−0.7	0.3	−2.2 (−36%)	−11.5	7.1
Stairs	0.7 (10%)	−3.6	5.1	0.1 (2%)	−3.8	4.1
Cycle	−0.5 (−2%)	−4.5	3.5	−0.6 (−2%)	−4.0	2.8

**Table 3.** Mean bias and limits of agreement (minutes) for Motus and ActiPASS comparisons. Mean bias and limits of agreement (LoA) in minutes for Bland-Altman comparisons of Motus (25 Hz) vs. ActiPASS (Fig. 5) and Motus (12.5 Hz) vs ActiPASS (Fig. 6). Relative values are based on the minutes classified by ActiPASS for each movement behaviour.

The Bland-Altman plot compares agreement between Motus (25 Hz) and ActiPASS in estimating minutes spent in each movement behaviour. The mean bias (mean difference) is presented, where a bias close to zero indicates good agreement. A positive bias indicates that Motus classified more minutes in a given behaviour than ActiPASS, while a negative bias suggests that ActiPASS classified more minutes. The dashed lines represent the 95% limits of agreement (mean difference  $\pm 1.96 \times$  standard deviation), with narrower limits indicating more consistent results and wider limits suggesting greater variability.

Figure 5 shows the agreement between Motus 25 Hz and ActiPASS, with a mean bias of  $\pm 1$  min for all behaviours. Relative to the total time classified for each specific behaviour, this corresponds to  $\pm 1\%$  of the time in sedentary, standing, moving and walking, while for the behaviours running, stair climbing and cycling, the mean bias results in  $-3\%$ ,  $10\%$  and  $-2\%$  respectively. The narrowest limits of agreement were observed for sedentary ( $-1.5$  to  $1.4$  min), while the greatest limits of agreement were observed for stair climbing ( $-3.6$  to  $5.1$  min).

The Bland-Altman plot compares agreement between Motus (12.5 Hz) and ActiPASS in estimating minutes spent in each movement behaviour. The mean bias (mean difference) is presented, where a bias close to zero indicates good agreement. A positive bias indicates that Motus classified more minutes in a given behaviour than ActiPASS, whereas a negative bias suggests that ActiPASS classified more minutes. The dashed lines represent the 95% limits of agreement (mean difference  $\pm 1.96 \times$  standard deviation), with narrower limits indicating more consistent results and wider limits suggesting greater variability.

Figure 6 shows the agreement between Motus 12.5 Hz and ActiPASS with a mean bias of  $\pm 1$  min for all behaviours except for standing, walking and running, with mean biases of  $5.1$  min,  $-2.9$  min and  $-2.2$  min,

respectively. Relative to the total time classified in each specific behaviour, this corresponds to 0%, 1%, 2% and -2% for sedentary, moving, stair climbing and cycling, while for standing, walking and running, this corresponds to 3%, -4% and -36%. For limits of agreement, the narrowest limits were observed for sedentary (-2.1 to 2.0 min), while the widest limits were seen for standing (-4.1 to 14.3 min).

No specific trend of the data spread was seen for any of the behaviours when comparing agreement between Motus 25 Hz and ActiPASS (Fig. 5) or Motus 12.5 Hz and ActiPASS (Fig. 6).

## Discussion

This study investigated whether Motus performance could be maintained while reducing its sampling frequency from 25 Hz to 12.5 Hz. Thus, we aimed to validate Motus against video observations under standard and semi-structured laboratory conditions, and to compare Motus with the ActiPASS tool in free-living conditions. The findings indicate that reducing the sampling frequency from 25 Hz to 12.5 Hz does not substantially compromise the performance of Motus. Additionally, both frequencies demonstrated strong agreement in performance (F1-score and balanced accuracy) and against ActiPASS across most behaviours.

These results align with previous studies<sup>16,20,23</sup> and further support the viability of low-frequency accelerometry for detecting basic movement behaviours in daily living. For example, Lendt et al.<sup>23</sup> reported balanced accuracies ranging from 0.82 to 0.99 for similar movement behaviours, using a sampling frequency as low as 12.5 Hz in both laboratory and free-living conditions. Similarly, Khusainow et al.<sup>20</sup>, tested six sampling frequencies (10–60 Hz) and different machine learning algorithms and found that the classifiers achieved meaningful classification performance (accuracy) for sampling frequency as low as 10 Hz ( $\geq 93\%$ ), with peak performance at 20 Hz for different movement behaviours, and no statistically significant relationship between sampling frequency and classification performance. Moreover, Kahn et al.<sup>16</sup> investigated optimal sampling frequencies, identifying non-optimal frequencies and emphasising the waste of resources, as higher sampling rates require more battery and storage usage. For example, their results showed that reducing sampling rates from much higher original values (e.g., 96 Hz for quality control activities in manufacturing, 100 Hz for physical activities such as running and cycling, and 250 Hz for walking and gait recognition) to lower ones, such as 32 Hz, 17 Hz, and 12 Hz, led to significant reductions in battery and storage consumption without sacrificing classification performance. Together, these studies highlight the suitability of low sampling frequencies for accelerometer-based classification of human movement behaviours. Additionally, they emphasise that lower sampling frequencies in large-scale surveillance can reduce data transfer time, conserve storage capacity, and extend accelerometer battery life.

Despite this, we did observe some differences between Motus 25 and 12.5 in classifying vigorous activity, such as running. Specifically, 12.5 Hz misclassified running more often than 25 Hz (95.8% vs. 99.8% respectively) in the laboratory, compared with video. We also identified wider limits of agreement for running compared with the other behaviours, suggesting that, for some participants, the difference in F1-score could be larger even though the overall agreement in F1-score was strong. Furthermore, in free-living, Motus 25 and ActiPASS were similar, while Motus 12.5 and ActiPASS showed some differences, with Motus 12.5 Hz classifying 2.2 min less for running compared with ActiPASS. This indicates that Motus 12.5 Hz underestimates running compared with ActiPASS. However, running is a relatively rare behaviour, accounting for 3.9 to 6.3 min of the whole day, resulting in a relatively large underestimation. This underestimation may influence downstream analyses, particularly in studies focusing on physical activity intensity patterns or adherence to guidelines for vigorous activity. As a potential mitigation strategy, we note that such vigorous activity could be combined with similar activity categories (e.g., collapsing running with brisk walking or general moderate to vigorous activity categories) in analytical models, depending on the study purpose.

The potential underestimation of vigorous activities such as running may be explained by two factors. First, running involves higher frequencies within the human movement range (0.5–10 Hz)<sup>45,46</sup>, and a sampling rate of 12.5 may be too low to capture these accurately without aliasing, as required by the Nyquist–Shannon sampling theorem<sup>16,47</sup>. Second, even if the frequency content is preserved, low sampling rates can miss short, high-amplitude events like foot strikes, leading to underestimation of an variability in the signal. While cadence tends to remain similar across different running speeds, stride length increases, resulting in greater impact forces when the feet hit the ground. Because ActiMotus uses the standard deviation of the signal from these impacts to identify running, the standard deviation may appear lower due to missing high-frequency and/or high-amplitude data, leading to misclassification at 12.5 Hz<sup>5</sup>. Ultimately, the choice of sampling frequency should be guided by the specific research question and the accepted trade-off between the need for data resolution and practical constraints such as battery life and storage capacity.

Besides looking at the differences between the two sampling frequencies, we observed stair climbing to be the movement behaviour with the poorest performance for both sampling frequencies. Stair climbing had the lowest overall performance (F1-score: 0.87 for 25 Hz and 0.86 for 12.5 Hz), indicating a less strong balance between precision and recall compared with the other movement behaviours ( $\geq 93\%$ ). However, the agreement between 12.5 Hz and 25 Hz, as well as ActiPASS, was strong, indicating no systematic under- or overestimations between sampling frequencies or systems. Still, the limits of agreement were wider compared with the other behaviours, indicating that, for some participants, the difference could be larger. This is also seen for similar studies<sup>6,33</sup> and can be explained by how the algorithm is developed. ActiMotus (the algorithm in Motus) uses a fixed decision tree and calculates an individual threshold for each participant, based on available data, to distinguish walking from stair climbing<sup>5</sup>. Typically, this threshold becomes more precise with larger datasets. However, in this study, each participant had only ~ 20 min of laboratory data, significantly less than the extensive data (e.g., seven days) used in surveillance studies, leading to greater uncertainty. This limited data volume increases uncertainty in threshold estimation, making misclassification more likely. Since stair climbing and walking have similar movement patterns, the key distinguishing factor is the angle of the thigh (sensor placement), specifically, how high the thigh is lifted. If the threshold for detecting stair climbing is set too high, the algorithm may misclassify

stair climbing as walking. This is supported by the data, which shows that approximately 8% of stair-climbing was incorrectly classified as walking. Future studies should investigate the sensitivity to the amount of data used for calculating thresholds, as well as how staircase height, sensor placement on the thigh (distance from the knee), and leg length influence classification accuracy.

### Strengths and limitations

One of the notable strengths of our study was the inclusion of two independent annotators, which enhanced both the reliability and validity of the annotated data used for validation reference. Additionally, the validation incorporated data from both structured and semi-structured settings, ensuring applicability to real-life scenarios, with video observations serving as the gold standard for validating movement behaviours. Furthermore, the two-step calibration procedure, including device-specific autocalibration<sup>31</sup> and individual-specific calibration<sup>36</sup>, minimised potential sensor- and individual-related variability. We also applied a combination of complementary evaluation methods, including confusion matrices, F1-scores, balanced accuracy, and Bland–Altman plots to capture different aspects of system performance. This multi-method approach allows for both straightforward comparisons with existing systems and a more detailed assessment of class-wise performance, systematic differences, and agreement across settings. Finally, the Motus system was compared with an already validated and widely used system in research, ActiPASS, which enhances the reliability and contextual relevance of the results<sup>5,6,33</sup>.

While the free-living comparison was not intended as a validation, the absence of video or other external ground truth still limits the certainty of behaviour classification under real-world conditions. However, capturing free-living activities in a natural environment is extremely time-consuming and presents challenges, not only due to ethical considerations and privacy concerns but also due to the complexity of annotation<sup>48</sup>. The annotation process, which relies on human observation to label activities, can inadvertently misinterpret certain behaviours, further complicating the validation of free-living data. This issue was particularly evident during cycling when the participant was not pedalling, where the accelerometer was positioned on the leg placed on the lowest pedal. Motus identified the accelerometer's position as “standing,” but the annotators interpreted it as “sedentary,” as they observed the behaviour instead of the accelerometer orientation. This discrepancy highlights the potential for human error in the annotation process, rather than issues with the algorithm itself<sup>48</sup>. While structured procedures were in place and inter-rater agreement was high, further improvements are needed to enhance annotation validity and reduce the risk of shared interpretation bias. In this study, annotators received initial training and followed written protocols, but future work could include a pilot annotation phase with iterative refinement of both procedures and instructional material<sup>49</sup>. This may improve clarity, especially around complex transitions and ambiguous behaviours. Additional accelerometer placements (e.g., on both thighs or the back) could help confirm postural orientation and reduce reliance on visual interpretation alone<sup>30</sup>. Finally, the development of automated annotation tools, potentially using machine learning or artificial intelligence, may also support more scalable and objective labelling in future validation studies.

In addition to annotation-related errors, participant instructions can also introduce misclassification. For instance, during the cycling behaviour, participants were allowed to adjust their seat height freely. However, one participant did not modify the seat after the previous user, resulting in a seat that was too high. Consequently, they were forced to sit on the front edge, altering their leg positioning. As a result, their leg did not reach the angle threshold set by the algorithm for detecting cycling, leading to misclassification. This could potentially explain why all the cycling behaviour for this participant was classified as stair climbing, resulting in poor performance for this participant (F1-score for cycling: 0.00 and 0.03 for 25 Hz and 12.5 Hz, respectively). This issue could have been minimised by ensuring that participants were properly instructed on the importance of a natural seating position. Checking their posture, confirming that they were seated comfortably, and allowing sufficient time for adjustments could have helped reduce such errors. However, it is also possible that such non-optimal cycling postures occur in real-life situations, suggesting that the observed misclassification may reflect a real-world challenge for the system rather than solely a protocol deviation.

### Practical implications and future development

The findings from the present study have important practical implications. Our results show that reducing the sampling frequency does not significantly affect the performance of Motus. This could enhance the overall feasibility of Motus, which has already been demonstrated to be feasible for large-scale data collection<sup>14,15</sup>. However, the feasibility study, conducted in two stages, identified issues in the first stage with time-consuming data transfer, referring to the time required for data to be transmitted from the accelerometer to the smartphone via Bluetooth and subsequent upload to cloud storage. In the second stage of the feasibility trial, when lowering the sampling frequency from 25 Hz to 12.5 Hz, the data transfer time was reduced by approximately 50%. Additionally, by lowering the sampling frequency, the accelerometer generates less data per unit of time, allowing it to store data for longer periods without needing to transfer it to the cloud, thus improving the system's practical applicability. This suggests that reducing the sampling frequency could be a viable strategy to improve data storage and data transfer efficiency, making the system more suitable for long-term monitoring in real-world applications.

Future development of the Motus system will focus on optimising the performance of Motus at 12.5 Hz and addressing the issues identified in this study. The primary focus is on high-intensity movement behaviours, such as running, as our findings indicate that these behaviours are most affected by reduced sampling frequency. As a result, we have begun refining further adjustments to the correctional polynomial used for the integration of the SENSmotionPlus accelerometer to account for the reduced sampling frequency. A result from this is presented in Additional File 4 and Figure A1. Moving forward, additional refinement of these adjustments, along

with testing across various populations, will be crucial. Such development will be open-source and available on GitHub<sup>26</sup> to facilitate ongoing collaboration within the research community.

## Conclusion

In conclusion, our study provides significant insights into the performance of the Motus system, highlighting the possibility of reducing the sampling frequency from 25 Hz to 12.5 Hz and still obtaining high performance in classifying different movement behaviours. These findings contribute to the existing literature and offer practical solutions for enhancing activity monitoring in real-world settings.

## Data availability

The data that support the findings of this study can be made available upon reasonable request by contacting the corresponding author at xtpl@nfa.dk. The processing tools ActiPASS ([<https://github.com/Ergo-Tools/ActiPASS>]) and ActiMotus ([<https://github.com/acti-motus/acti-motus>]) are publicly available at GitHub.

Received: 10 March 2025; Accepted: 28 October 2025

Published online: 27 November 2025

## References

1. World Health Organization. *WHO Guidelines on Physical Activity and Sedentary Behaviour* (World Health Organization, 2020).
2. Stevens, M. L. et al. Thigh-worn accelerometry for measuring movement and posture across the 24-hour cycle: a scoping review and expert statement. *BMJ open. Sport Exerc. Med.* **6** (1), e000874 (2020).
3. Stamatakis, E. et al. Emerging collaborative research platforms for the next generation of physical activity, sleep and exercise medicine guidelines: the prospective physical Activity, Sitting, and sleep consortium (ProPASS). *Br. J. Sports Med.* **54** (8), 435–437 (2020).
4. Hettiarachchi, P. & Johansson, P. A. P. A. S. S. Available from: <https://github.com/Ergo-Tools/ActiPASS> (2024).
5. Skotte, J., Korshøj, M., Kristiansen, J., Hanisch, C. & Holtermann, A. Detection of physical activity types using triaxial accelerometers. *J. Phys. Act. Health.* **11** (1), 76–84 (2014).
6. Stemland, I. et al. Validity of the Acti4 method for detection of physical activity types in free-living settings: comparison with video analysis. *Ergonomics* **58** (6), 953–965 (2015).
7. Crowley, P. et al. Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: a proof-of-concept for the prospective physical Activity, Sitting, and sleep consortium (ProPASS). *Int. J. Behav. Nutr. Phys. Activity.* **16** (1), 65 (2019).
8. Evenson, K. R., Scherer, E., Peter, K. M., Cuthbertson, C. C. & Eckman, S. Historical development of accelerometry measures and methods for physical activity and sedentary behavior research worldwide: A scoping review of observational studies of adults. *PloS One.* **17** (11), e0276890 (2022).
9. Gupta, N. et al. Are we ready for wearable-based global physical activity surveillance? *Br. J. Sports Med.* **58** (7), 356–358 (2024).
10. Lee, P. H., Macfarlane, D. J. & Lam, T. H. Factors associated with participant compliance in studies using accelerometers. *Gait Posture.* **38** (4), 912–917 (2013).
11. Christiansen, L. B. et al. Device-based physical activity measures for population surveillance-issues of selection bias and reactivity. *Front. Sports Act. Living.* **5**, 1236870 (2023).
12. Crowley, P. et al. A novel system for the Device-Based measurement of physical Activity, sedentary Behavior, and sleep (Motus): usability evaluation. *JMIR Formative Res.* **7**, e48209 (2023).
13. Crowley, P. et al. The surveillance of physical Activity, sedentary Behavior, and sleep: protocol for the development and feasibility evaluation of a novel measurement system. *JMIR Res. Protocols.* **11** (6), e35697 (2022).
14. Gupta, N. et al. A feasibility study of the “Motus” system for wearable-based movement behaviors at scale. *J. Phys. Act. Health.* **22**(10), 1244–1255 (2025).
15. Eghøj, M. et al. Fysisk aktivitet, stillesiddende adfærd og søvn: Resultater fra monitorering med accelerometre i Danskernes sundhed 2023. (2024).
16. Khan, A., Hammerla, N., Mellor, S. & Plöts, T. Optimising sampling rates for accelerometer-based human activity recognition. *Pattern Recognit. Lett.* **73**, 33–40 (2016).
17. Migueles, J. H. et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Med. (Auckland NZ).* **47** (9), 1821–1845 (2017).
18. Junker, H., Lukowicz, P. & Troster, G. (eds) Sampling frequency, signal resolution and the accuracy of wearable context recognition systems. Eighth International Symposium on Wearable Computers; 31 Oct.-3 Nov. 2004. (2004).
19. Yang, X., Dinh, A. & Chen, L. (eds) Implementation of a wearable real-time system for physical activity recognition based on Naive Bayes classifier. 2010 International Conference on Bioinformatics and Biomedical Technology; 16–18 April 2010. (2010).
20. Khusainov, R., Azzi, D., Achumba, I. E. & Bersch, S. D. Real-Time human Ambulation, Activity, and physiological monitoring: taxonomy of Issues, Techniques, applications. *Challenges Limitations.* **13** (10), 12852–12902 (2013).
21. Shannon, C. E. Communication in the Presence of Noise. *Proceedings of the IRE.* **37**(1), 10–21 (1949).
22. Jagos, H., Oberzaucher, J., Reichel, M., Zagler, W. L. & Hlauschek, W. A multimodal approach for insole motion measurement and analysis. *Procedia Eng.* **2** (2), 3103–3108 (2010).
23. Lendt, C., Braun, T., Biallas, B., Froböse, I. & Johansson, P. J. Thigh-worn accelerometry: a comparative study of two no-code classification methods for identifying physical activity types. *Int. J. Behav. Nutr. Phys. Activity.* **21** (1), 77 (2024).
24. Bartholdy, C. et al. Reliability and construct validity of the SENS Motion<sup>®</sup> activity measurement system as a tool to detect sedentary behaviour in patients with knee osteoarthritis. *Arthritis* **2018**, 6596278 (2018).
25. Pedersen, B. S. et al. Validation of two activity monitors in slow and fast walking hospitalized patients. *Rehabilitation Res. Pract.* **2022**, 9230081 (2022).
26. The National Research Centre for the Working Environment (NEA). ActiMotus. Available from: <https://github.com/acti-motus/acti-motus> (2024).
27. Korshøj, M. et al. Validity of the Acti4 software using actigraph GT3X+ accelerometer for recording of arm and upper body inclination in simulated work tasks. *Ergonomics* **57** (2), 247–253 (2014).
28. Ingebrigtsen, J. et al. Validation of a commercial and custom made Accelerometer-Based software for step count and frequency during walking and running. *J. Ergon.* **03**, 02 (2013).
29. Johansson, P. J. et al. Development and performance of a sleep Estimation algorithm using a single accelerometer placed on the thigh: an evaluation against polysomnography. *J. Sleep. Res.* **32**(2), e13725 (2022).
30. Hettiarachchi, P. et al. Validity of a Non-Proprietary algorithm for identifying lying down using Raw data from Thigh-Worn. *Triaxial Accelerometers.* **21** (3), 904 (2021).

31. van Hees, V. T. et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *Journal of applied physiology* (Bethesda, Md: 2014;117(7):738–44. (1985).
32. Grant, P. M., Ryan, C. G., Tigbe, W. W. & Granat, M. H. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *Br. J. Sports Med.* **40** (12), 992–997 (2006).
33. Logacjov, A., Bach, K., Kongsvold, A., Bårdstu, H. B. & Mork, P. J. HARTH: A human activity recognition dataset for machine learning. *Sensors* **21**(23), 7853 (2021).
34. Kipp MJHocp. Anvil: The video annotation research tool. :420–36. (2014).
35. McHugh, M. L. Interrater reliability: the kappa statistic. *Biochem. Med. (Zagreb)*. **22** (3), 276–282 (2012).
36. Højvall, C. Detection of physical behavior from thigh worn accelerometer: validation of a new data processing software that automatically compensates for minor variations in the placement of the accelerometer. MSc thesis, KTH Royal Institute of Technology (2020).
37. Brønd, J. C., Pedersen, N. H., Larsen, K. T. & Grøntved, A. Temporal alignment of dual monitor accelerometry recordings. *Sensors* **21**(14), 4777 (2021).
38. Nazare, A. C., Costa, F. & Schwartz, W. R. (eds) Content-Based Multi-Camera Video Alignment using Accelerometer Data. 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS); 2018 27–30 Nov. 2018. (2018).
39. Scarpari, J. R. S. et al. (eds) Method for the Synchronization of Data Recorders by Coupling Accelerometer Data. IEEE International Symposium on Inertial Sensors and Systems (INERTIAL); 2021 22–25 March 2021. (2021).
40. McKinney, W. Data structures for statistical computing in python. *SciPy* **445** (1), 51–56 (2010).
41. Harris, C. R. et al. Array programming with numpy. *Nature* **585** (7825), 357–362 (2020).
42. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods*. **17** (3), 261–272 (2020).
43. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
44. Altman, D. G. & Bland, J. M. Measurement in medicine: the analysis of method comparison studies. *J. Royal Stat. Society: Ser. D (The Statistician)*. **32** (3), 307–317 (1983).
45. Brønd, J. C., Andersen, L. B. & Arvidsson, D. Generating actigraph counts from Raw acceleration recorded by an alternative monitor. *Med. Sci. Sports. Exerc.* **49** (11), 2351–2360 (2017).
46. Fridolfsson, J., Börjesson, M. & Arvidsson, D. A Biomechanical Re-Examination of physical activity measurement with accelerometers. *Sensors* **18** (10), 3399 (2018).
47. Jerri, A. J. The Shannon sampling theorem—Its various extensions and applications: A tutorial review. *Proceedings of the IEEE*. **65**(11), 1565–96 (1977).
48. Jones, M., Byun, C., Johnson, N. & Seppi, K. Understanding the roles of video and sensor data in the annotation of human activities. *Int. J. Human–Computer Interact.* **39** (18), 3634–3648 (2023).
49. Bourke, A. K., Ihlen, E. A. F. & Helbostad, J. L. Development of a gold-standard method for the identification of sedentary, light and moderate physical activities in older adults: definitions for video annotation. *J. Sci. Med. Sport.* **22** (5), 557–561 (2019).

## Acknowledgements

The authors would like to thank all study participants for their time and effort. We also thank everyone who assisted with data collection for their valuable contributions.

## Author contributions

TPL designed and coordinated the study activities and was the primary author of the manuscript. SASH and JRL conducted data analysis and contributed to methodology development. CBL was responsible for designing and conducting the data collection. PH provided technical support and contributed to data interpretation. MB supervised the data collection process, provided methodological advice, and assisted in interpreting the results. NG supervised the overall project and guided throughout. All authors provided feedback, critically reviewed, and approved the final manuscript.

## Funding

The study was funded by the Danish Work Environment Research Fund (AMFF; grant number 20235100781).

## Declarations

## Competing interests

The authors declare no competing interests.

## Ethics approval and consent to participate

In accordance with the Helsinki Declaration, participants received written and oral study information before providing written informed consent. The study was verified by The Danish Ethics Committee (j.nr. F-24042892) to be approved via the National Research Centre for the Working Environments authorisation for low-risk non-invasive studies on healthy consenting adults.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-26373-7>.

**Correspondence** and requests for materials should be addressed to T.P.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025