Accuracy of elevation recording using sport watches while walking and running on hilly and flat terrain

Rahel Ammann[1,2] · **Wolfgang Taube**[2] · **Fabian Kummer**[1] · **Thomas Wyss**[1]

Abstract Elevation gain (EG) is a significant contributor to the total workload in many endurance sports. Hence, the aim of this study was to evaluate the accuracy of elevation recording as assessed by popular sport watches. Eighteen participants walked and ran at different speeds in various weather conditions in two terrain types: on a hilly 2490 m course with a total EG of 90 m and on a flat 1200 m outdoor track with 0 m EG. In total, 180 recordings from each sport watch were analyzed and compared according to two processing types: filtered and unfiltered EG data. Compared to the reference values, regarding default settings, on hilly terrain, EG was underestimated by -3.3 to -9.8 %, and on flat terrain, EG was overestimated by 0.0–4.8 m per 1200 m. These errors could be reduced to -3.3 to $+0.4$ %, when filtering conditions were adjusted according to the terrain. Gait speed (ranging from 1.47 to 4.89 m s^{-1}) or fluctuations in weather conditions between- or within-trials did not influence EG accuracy. A straightforward comparison between manufacturers is hampered as the filter conditions set by default differ. In conclusion, all devices measure EG adequately; however, the displayed default feedback on EG data is not always the most accurate measurement. Consequently, accuracy in elevation recordings could be increased if users appropriately post-process EG data.

✉ Rahel Ammann
 Rahel.Ammann@baspo.admin.ch

1 Swiss Federal Institute of Sport Magglingen SFISM, Magglingen, Switzerland

2 Department of Medicine, Movement and Sport Science, University of Fribourg, Fribourg, Switzerland

1 Introduction

Endurance athletes, coaches, and researchers rely on objectively assessed measurements to quantify, monitor, and evaluate physiological training workloads [1, 2]. Variables such as time, speed, distance, heart rate, and elevation gain (EG) are collected during most workouts and competitions. The information about EG is important in endurance sports, such as orienteering, mountain running, mountain biking, or cross-country skiing, because it contributes greatly to the total performance time, and consequently, to workload [3, 4]. A ratio of 1 to 10 and 1 to 8 for women and men, respectively, has previously been published to compare the exercise time of vertical distances covered on foot with the exercise time of horizontal distances [5]. This means that the time used for a 100 and 125 m incline for women and men, respectively, equals the time used for a 1000 m flat distance. Hence, especially in endurance sports with a high level of incline, technology that provides feedback about EG is important to accurately monitor and analyze workload during training and competition.

A range of wearable measurement systems, such as sport watches or small computers, is available on the market for monitoring EG during outdoor activities. Menaspà and colleagues [6] investigated the consistency of commercial devices for measuring EG. They demonstrated that measures of EG were relatively consistent within each manufacturer. Yet, they did not assess the absolute measurement error. Therefore, concurrent validity of the devices for measuring EG has not been investigated, so far. Hence, the purpose of this study was to validate three popular sport

watches in terms of the EG recorded during different gait speeds and various weather conditions on hilly and flat terrains. Furthermore, the influence of processing types, i.e., filtered and unfiltered data, on EG errors was investigated.

2 Materials and methods

2.1 Participants

Ten healthy male and eight healthy female volunteers (24.7 ± 1.8 years, 175.7 ± 7.3 cm, 69.5 ± 8.4 kg) participated in this study. Informed consent was obtained from all participants after familiarization with the procedure. The research procedure was approved by the Institutional Review Board of the Federal Office of Sport.

2.2 Methodology

The investigated devices were the Garmin® Forerunner 910XT (GF; Garmin® Ltd., Southampton, UK), the Polar® RS800CX (PRS; Polar® Electro Oy, Kempele, Finland), and the Suunto® Ambit2 (SA; Suunto®, Valimotie, Finland). The GF and the SA calculate elevation using a barometric altimeter corrected by GPS data, whereas the PRS has a built-in barometer only. However, the manufacturers do not provide detailed information about the underlying mechanisms of how EG is derived. Furthermore, a stopwatch (Stopstart 2, Hanhart, Hamburg, Germany) and an air pressure gauge (GDH 200-14, GHM Messtechnik, Regenstauf, Germany) were included to control for the influence of speed and weather conditions, respectively, on EG errors.

The study group completed a total of 180 trials on two different courses, each of which was free of any trees or buildings. Participants completed each course at three speeds introduced as representative of individual normal walking, slow running, and fast running. As many sports trails with inclines also include flat sections, it is important to evaluate whether the devices measure no EG on flat terrain. Therefore, the first course consisted of three rounds on a 400-m outdoor synthetic track, resulting in 1200 m of flat distance. The second course was 2490 m with a total EG of 90 m [7]. To account for natural training and competition settings, in practice with predominantly intermittent ascent and descent, no course with continuous ascent only was chosen. Hence, the hilly course was composed of a 415-m asphalt track with an EG of 30 m (average slope 7.2 %), which was completed three times up and down; thus, the start and endpoint were at the same altitude. Every time the participant passed the marked turning points at the highest and lowest altitude, they stood still for 2 s before

resuming the course. This was done to ensure that participants, and consequently, the watches, really reached the altitude level of the reference. Each test day, the investigated devices were calibrated to the baseline elevation of 857 m above sea level. The epoch duration in all devices was set to the highest possible value, equal to 1 s. Sport watches from all three manufacturers were worn simultaneously, placed randomly on either the right or left wrist. Data were uploaded to the corresponding webpage or software (Garmin® Connect, Polar® ProTrainer 5, and Movescount, respectively). The measurements were collected over a period of 3 months to account for differing weather conditions and times of day. Speed was averaged over the entire distance of one trial. Fluctuations in weather conditions between- and within-trials were derived from air pressure values noted before and after each course.

2.3 Data analysis

Two different data sets from each sport watch were evaluated; filtered and unfiltered EG data, hereafter referred to as processing type. All manufacturers have the option of exporting the raw (unfiltered) data as well as data filtered with an incorporated filter in the firmware of their devices or in their computer software/webpages that corrects the elevation recordings. We cannot make any statements about the filter settings, as none of the manufactures provided such information upon enquiry. Importantly, the processing type, set by default, is not the same for all three manufacturers. In a first step, we analyzed the data from the default setting within each device, as users rely on this data for feedback about EG. Within the PRS and the SA, the filtered EG data are always displayed by default. In contrast, the GF provides unfiltered EG data by default. In a second analysis, we investigated the data not set by default. This means that within the PRS and the SA, we used the unfiltered EG data and exported them to Microsoft Excel (2011) to manually analyze EG data in 1-s epochs. Every change in EG was added to a total EG value. Instead, the Garmin service provides an elevation correction option on the webpage that post-processes the EG data.

2.4 Statistical analysis

All results were expressed as absolute means ± SD, absolute differences ± SD to the reference, and relative differences ± SD to the reference. The latter was expressed twofold: the recorded EG from the hilly course was divided by 90 m multiplied by 100. However, as the reference value on the flat course was 0 m, the usual relative difference calculation and other statistical approaches could not be applied. Therefore, the fictive unit percentage EG meters per 1200 m (entire distance of the flat course)

was used to calculate relative difference. The within-device variability was calculated using coefficients of variation (CVs) for the data collected on hilly terrain. Normal distribution was assumed when the ratio of skewness to the standard deviation of skewness did not exceed ± 2.0. The characteristics of absolute EG differences to the reference were illustrated by means of boxplots. Linear regression analysis was performed to evaluate whether device type, terrain type (hilly vs. flat), processing type (filtered vs. unfiltered EG data), speed, or fluctuations in weather conditions between- and within-trials were predictors of measurement errors in EG, using relative differences as a dependent variable. Relative differences in EG were compared between the manufacturers using one-way ANOVA. In case of significant effects, a Bonferroni post hoc analysis was conducted. The statistical analyses were executed with IBM SPSS Statistics 22 (IBM Corporation, Armonk, NY, USA), and significance was set at $P \leq 0.05$.

3 Results

When considering the default settings, i.e., unfiltered EG data within the GF and filtered EG data within the PRS and the SA, on the hilly course with a 90-m incline, the mean EG data were 87.0 ± 5.0, 81.2 ± 3.1, and 84.4 ± 4.7 m as recorded by the GF, the PRS, and the SA, respectively (Table 1a). Consequently, EG was underestimated by -3.3, -9.8, and -6.2 %, respectively. On the flat course with a 0-m incline, the mean EG data showed 4.8 ± 5.6, 0 ± 0, and 0.3 ± 1.3 m, which were relative overestimations per 1200 m horizontal distance of 0.4, 0.0, and 0.03 % by the GF, the PRS, and the SA, respectively (Table 1b). The CVs for the GF, the PRS, and the SA were 5.8, 3.8, and 5.5 %, respectively. Regression analysis revealed that the device type, terrain, and processing type were significant (P values <0.001) predictors of enlarged relative errors in EG. On the hilly terrain, the basic underestimation of EG was significantly reduced by around 12 % in the unfiltered data compared to the filtered EG data in all sport watches (-1.7 ± 4.5 vs. -13.4 ± 8.9 %; Table 1a). By contrast, on the flat course, for the PRS and the SA, EG errors only occurred in the unfiltered data in comparison to the filtered EG data (0.5 ± 0.3 vs. 0.0 ± 0.1 %; Table 1b). However, for the GF, again the unfiltered EG data were more accurate than the filtered EG data (0.4 ± 0.5 vs. 1.7 ± 0.1 %). The diverging impact of the processing type and terrain type on EG errors by each manufacturer is depicted in Fig. 1. When analyzing the sum of the EG data from the hilly and flat section, significant between-device variability occurred ($F_{2,23.899} = 21.21$, $P < 0.001$). The Bonferroni post hoc corrections revealed significant differences between all devices with overall differences of -1.4, -4.8, and -3.1 %

in comparison to the references within the GF, the PRS, and the SA, respectively. Generally, speed and weather conditions (air pressure changes between- and within-trials) were not significant predictors of EG errors. Gait speed differed significantly ($F_{2,1035} = 1564.23$, $P < 0.001$) between normal walking, slow running, and fast running, with speeds of 2.02 ± 0.29, 2.98 ± 0.41, and 3.85 ± 0.55 m s^{-1}, respectively. Overall, the between-trial air pressure was 912.05 ± 10.43 hPa, and changes in air pressure within-trials were -0.15 ± 0.58 hPa. The 180 trials lasted on average 16 min 40 s $\pm$ 5 min 00 s.

4 Discussion

Elevation gain is a relevant contributor to athletes' total workload. Therefore, the present study investigated how accurately popular sport watches record EG. In general, considering the default settings on hilly terrain, EG was underestimated by -3.3 to -9.8 %, while on flat terrain, EG tended to be overestimated (0.0–0.4 %). This knowledge is relevant to the accurate interpretation of data output from training sessions and competitions on either predominantly flat or hilly terrain in terms of quantification, monitoring, and evaluation of an activity.

Possible reasons for errors in EG estimations include gait-related arm swing, varying quality in GPS signal connections, and/or fluctuation in air pressure [2, 8]. The latter could not be confirmed in the present study. Yet, each trial lasted only approximately 15 min, and thus, within-trial fluctuations in air pressure may still be an issue during activities of longer duration. These aforementioned reasons may cause the devices to register EG, even though the terrain is flat. To minimize this effect, the PRS and the SA incorporate data correction algorithms as default settings that remove small fluctuations from the cumulative EG. Accordingly, the recorded differences in EG on the flat terrain were negligible with 0.0 and 0.2 m per 1200 m in the PRS and the SA, respectively, when considering filtered EG data. The unfiltered EG data from these devices revealed larger errors of 4.4 and 7.1 m, respectively. The GF offers a data correction algorithm as well, but not as a default setting. The recommended unfiltered EG data showed 4.8 m EG per 1200 m, and the filtered EG data returned 20.1 m EG on the flat terrain. None of the manufacturers provided their correction settings upon enquiry.

On hilly terrain, EG was underestimated. In addition to the previous explanations, it may be that the positive and negative peaks of elevation points are cut off due to low resolution rates. This implies that the full amplitude might not be recorded, because each low-elevation point is slightly increased and each high-elevation point is slightly reduced. In the present study, the frequencies for data

Table 1 Elevation gain recorded by three sport watches on a hilly course with a 90-m incline over a distance of 2490 m (a) and on a flat course with a 0-m incline over the distance of 1200 m (b)

(a) Sport watch	Filtered data			Unfiltered data		
	EG (m)	Mean difference to 90 m (m)	Relative difference to 90 m (%)	EG (m)	Mean difference to 90 m (m)	Relative difference to 90 m (%)
Garmin® Forerunner 910XT	68.0 ± 3.2	−22.0 ± 3.2	−24.4 ± 3.6	**87.0 ± 5.0**	**−3.0 ± 5.0**	**−3.3 ± 5.5**
Polar® RS800CX	**81.2 ± 3.1**	**−8.8 ± 3.1**	**−9.8 ± 3.5**	88.0 ± 1.8	−2.0 ± 1.8	−2.2 ± 2.0
Suunto® Ambit2	**84.4 ± 4.7**	**−5.6 ± 4.7**	**−6.2 ± 5.2**	90.3 ± 3.9	0.3 ± 3.9	0.3 ± 4.4
(b) Sport watch	Filtered data			Unfiltered data		
	EG (m)	Mean difference to 0 m (m)	Relative difference* (%)	EG (m)	Mean difference to 0 m (m)	Relative difference* (%)
Garmin® Forerunner 910XT	20.1 ± 1.2	20.1 ± 1.2	1.7 ± 0.1	**4.8 ± 5.6**	**4.8 ± 5.6**	**0.4 ± 0.5**
Polar® RS800CX	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	4.4 ± 3.1	4.4 ± 3.1	0.4 ± 0.3
Suunto® Ambit2	**0.2 ± 0.9**	**0.2 ± 0.9**	**0.03 ± 0.1**	7.1 ± 3.3	7.1 ± 3.3	0.6 ± 0.3

Data representing default setting in bold

Filtered data processed elevation gain data (default setting for the Polar and the Suunto watches), *unfiltered data* 1-s epoch elevation gain data (default setting for the Garmin watch), *EG* elevation gain
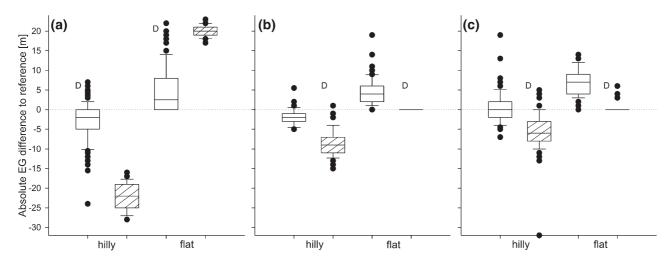
* Percentage EG meters per 1200-m flat distance

Fig. 1 Absolute elevation gain (EG) differences to the reference depending on processing type according to hilly (90 m EG) and flat (0 m EG) terrain. *Blank boxes* unfiltered EG data, *Dashed boxes* filtered EG data, *Solid horizontal line* median, **a** Garmin® Forerunner 910XT, **b** Polar® RS800CX, **c** Suunto® Ambit2, *D* default setting within that sport watch

collection were set to the highest possible value, equal to 1-s epoch time for each device. Moreover, to ensure standardized conditions, such as subjects really reaching the reference altitude, and to circumvent possible low-resolution issues, all participants remained still for 2 s at every turning point of the hilly course. Therefore, the underestimation of EG might be even larger under normal conditions. Interestingly, in all three sport watches, the underestimation was more pronounced, by around 12 %, in the filtered compared to the unfiltered EG data. This implies that the data correction algorithms, which are set by default within the PRS and the SA, are probably only appropriate on flat terrain. For the GF, the unfiltered EG data are set by default, and these data were more accurate on hilly terrain. Therefore, contrary to the flat course, the GF showed the most accurate EG of the three sport watches on the hilly terrain with a relative underestimation of the true value of −3.3 %, whereas the PRS and the SA reported underestimations of −9.8 and −6.4 %, respectively. The underestimation could be reduced to −2.2 and +0.3 % for the PRS and the SA, respectively, if considering unfiltered EG data.

Generally, comparability between devices is limited. The primary relevant difference between devices is the underlying processing type. Within the GF, the unfiltered (default setting) elevation recordings were always more accurate than the filtered elevation recordings. That is, the GF provides the user continuously with unfiltered EG data, always visible on the watch or at a glance on the output of the webpage. In contrast, the default filtered EG data from the PRS and the SA were beneficial on the flat terrain, but the filters were not appropriate for the hilly terrain, where a larger underestimation of the true EG occurred in the filtered compared to the unfiltered data. Thus, what seems to be an advantage on flat terrain turns out to be a disadvantage in terrains with incline. Unfortunately, the user cannot change the default settings in the PRS and the SA to obtain unfiltered data. This is particularly disturbing as it is a very time-consuming process to access the unfiltered EG data. When comparing the sport watches by their default EG recordings, the GF was the most accurate device when combining the data from both terrain types. However, as shown in Fig. 1 and by the CVs, the GF recorded the largest variations and outliers in EG differences for both terrains when compared to the PRS and the SA. Moreover, the estimated error in EG is influenced by the course profile, as the sport watches do not react in the same manner on hilly and flat terrains. In other words, when adding up the results for both terrain types, the length of each section may balance out the relative underestimation and overestimation of EG or may cause the total EG error to increase. In the present study, the hilly course was about twice the length of the flat course.

Our findings confirm and extend the results of Menaspà and colleagues [6], who evaluated device variability in the detection of EG between Garmin and SRM devices when assessed on the rooftop of a car and on the handlebars of bicycles. They observed deviations between brands and between different settings, i.e., filtered versus unfiltered data, similar to those found in the current study. However, as that previous study did not include reference to the actual EG, they could not assess the accuracy of the devices or clarify which setting is preferential and should be recommended. In the present study, EG data obtained during gait activities and processed twofold, were compared to an exact altitude reference. Furthermore, the hilly course represented a practical training setting of, e.g., trail runners or cross-country skiers, with intermittent ascents and descents. Yet, as the applied methodology attempted to reproduce ideal conditions, namely no trees or buildings and stops at turning points on the hilly course, one may assume that errors in EG may be larger under real-life conditions. Hence, further studies investigating these aspects are recommended.

5 Conclusion

The present study demonstrated that the commercially available sport watches were fairly accurate for assessment of EG while walking and running. However, considering the displayed data set as a default, on hilly terrain, the EG was underestimated by all three manufacturers (-3 to -10 %), while on flat terrain, the EG tended to be slightly overestimated ($<+1$ %). The GF provided more accurate results when considering the unfiltered EG data, which is set by default in that device. In contrast, for the PRS and the SA, having an active filter on flat terrains and an inactive filter on hilly terrains would theoretically be the optimal solution. However, this cannot be practically implemented by the user, as only filtered EG data are displayed by default. Consequently, the development of dynamic or adaptive filters that depend on the terrain might be an avenue of exploration for the manufacturers to further increase measurement accuracy in sport watches recording EG. In the meantime, one may post-process EG data manually to ensure the highest possible accuracy in the evaluation of total workload.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Fleming P, Young C, Dixon S, Carré M (2010) Athlete and coach perceptions of technology needs for evaluating running performance. Sports Eng 13:1–18
2. Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D (2011) Use of global positioning systems to study physical activity and the environment: a systematic review. Am J Prev Med 41:508–515
3. Minetti AE, Moia C, Roi GS, Susta D, Ferretti G (2002) Energy cost of walking and running at extreme uphill and downhill slopes. J Appl Physiol 93:1039–1046
4. Pivarnik JM, Sherman NW (1990) Responses of aerobically fit men and women to uphill/downhill walking and slow jogging. Med Sci Sports Exerc 22:127–130
5. Scarf P (2007) Route choice in mountain navigation, Naismith's rule, and the equivalence of distance and climb. J Sports Sci 25:719–726
6. Menaspà P, Impellizzeri FM, Haakonssen EC, Martin DT, Abbiss CR (2014) Consistency of commercial devices for measuring EG. Int J Sports Physiol Perform 9:884–886
7. Schwarz N (2013) BASPO Situation [map], scale 1:1,000, indicated objects with elevation accuracy of ±1 mm. Geoplan Hutzli+Klauser, Nidau
8. Parviainen J, Kantola J, Collin J (2008) Differential barometry in personal navigation. In: Position, location and navigation symposium, IEEE/ION, Monterey, CA, USA, 5–8 May 2008