2025 15th IEEE International Conference on Pattern Recognition Systems (ICPRS) | 979-8-3315-9170-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICPRS66293.2025.11302827

High-Precision Mapping of Mountain Terrain for Human Locomotion Research: A Multimodal Geospatial Framework

Raimundo Sanchez
School of Health &
Rehabilitation Sciences
The University of Queensland
Brisbane, Australia
Metric Trails
Brisbane, Australia
0000-0003-2242-1335

Claudio Nieto-Jimenez
Carrera de Entrenador Deportivo,
Facultad de Educación y
Ciencias Sociales
Universidad Andrés Bello
Viña del Mar, Chile
0000-0003-3017-2404

Cristian Riveros-Matthey
School of Health &
Rehabilitation Sciences
The University of Queensland
Brisbane, Australia
0000-0002-4341-3433

***Abstract*—Reliable trajectory data are essential for biomechanical and physiological studies conducted on variable outdoor terrain. However, consumer-grade GNSS devices and synthetic route planners often introduce spatial and vertical inaccuracies that affect derived metrics. This study presents a reproducible workflow for sub-meter trail mapping, tested on a 1 km mountain trail segment on Mt Coot-tha, Australia. Five consumer GNSS loggers (Garmin GPSMAP 67i, Coros Apex Pro II, Garmin Fenix 5S, Garmin Vivoactive 5, iPhone 12 mini) were recorded over six laps, then processed with resampling methods and elevation correction approaches, including elevation imputation from lidar-derived elevation models. Route-builder outputs from popular online platforms served as synthetic benchmarks. The dual-band GPSMAP 67i achieved maximum and mean 3D separation of 2.7 m and 0.6 m respectively when filtered by signal quality, outperforming any other scenario. Sub-meter trail mapping is achievable with consumer dual-band GNSS, RINEX data and detailed elevation data, while post-processing cannot compensate for poor horizontal fixes. Virtually designed routes should not replace high-quality empirical measurements in human locomotion studies.**

***Index Terms*—Global positioning system, Wearable sensors, Data fusion, Geospatial analysis, Light detection and ranging**

I. Introduction

Real-world locomotion research is indispensable. Mountain terrain offers locomotor challenges that no treadmill or level track can reproduce. Gradients shift from meter to meter, surfaces alternate between scree, wet roots and hardpan, and exposure changes with every switchback. These terrain variabilities create what some authors describe as locomotor complexity [1], a dynamic interaction between terrain features and motor control demands that affects pacing, gait transitions, and fatigue accumulation in ways that controlled platforms deliberately suppress. Metrics such as sustained gradient, terrain roughness, and curvature may also act as proxies for eccentric loading or neuromuscular strain, enabling terrain-level estimations of physiological cost [2]. Terrain variability elicits measurable changes in gait mechanics and cardio-respiratory responses during locomotion. Even a two percent change in slope shortens stride length, raises oxygen uptake and alters muscle activation patterns [3]. Field observations suggest that small terrain changes trigger gait transitions and modulate perceived exertion [4], [5]. Variables such as instantaneous speed, local gradient, accumulated elevation gain, and surface irregularity interact with individual characteristics, such as aerobic capacity, technical proficiency, and pacing strategy, to influence performance and potentially contribute to injury risk [6], [7]. To properly quantify these terrain-induced effects in real-world settings, accurate sensor-based measurement of both locomotor dynamics and environmental context is essential. While foot pods and inertial measurement units (IMUs) can reliably capture cadence, contact time, and segmental motion, they cannot accurately infer mechanical workload without precise terrain data, including slope and elevation profile [8]. Furthermore, consumer-grade GNSS devices are prone to signal drift under canopy cover or near cliffs; smartphone-based positioning often amplifies horizontal error. Without standardized spatial resolution, the complex topography of natural trails can distort estimates of distance and cumulative climb, thereby limiting biomechanical and physiological inference [9], [10]. Coarse Digital Elevation Models (DEMs) can introduce vertical errors exceeding 5 m in complex terrain [11]–[14]. Even high-end sport watches exhibit discrepancies of 3 - 9% in distance estimation, with even greater divergence in total ascent calculations [15], [16]. Because on-device smoothing and elevation correction are typically proprietary, derived metrics, such as grade-adjusted pace and elevation gain, may inherit hidden errors [16]. Studies comparing physiological effort across hills or slopes should explicitly report how distance and elevation gain were derived, corrected, and validated. True three-dimensional distance accounts for surface undulations and terrain features, unlike the geodesic two-dimensional chord length typically reported by GPS devices [17]. This discrepancy constrains direct comparisons between treadmill-based experiments (true 3D distance) and field studies, which often rely on simplified geodetic approximations. Repeated GNSS surveys, high resolution lidar scans, validated DEMs and open-positioning algorithms collectively help constrain spatial uncertainty across all three dimensions [18]–[20]. Relying on commercial or crowdsourced metrics without independent validation introduces substantial risk of systematic error. Crowd-sourced platforms such as Strava or Garmin Connect often publish tracks that have been simplified, elevation-corrected using DEMs of unknown resolution, and processed through undisclosed filtering algorithms. These layered inaccuracies cascade into leaderboard rankings, pacing plans and meta-analyses. Validation studies have reported consistent offsets in speed, distance and cadence when compared against reference systems [8]. In the absence of auditable and reproducible course measurement standards, key metrics, such as speed, gradient or mechanical power, cannot be reliably compared across athletes, events or studies. Recent efforts to address these measurement limitations generally follow two main approaches. Large-scale analyses extract aggregate patterns from millions of GNSS traces [21], [22]; however, the averaging inherent in such methods tends to obscure the fine-grained terrain variations that influence physiological cost. Field-based studies, by contrast, map short loops [23], [24] or entire trails with high-precision GNSS or lidar systems [18], achieving centimeter-scale accuracy but requiring specialized hardware and expertise. Other recent approaches have begun to integrate barometric altimetry with differential GNSS, comparing DEM resolutions across biomes, and experimenting with context-aware normalization techniques [6], [25], [26]. Yet despite these advances, a compact, openly documented, and auditable pipeline remains absent. This study aims to develop a transparent and replicable framework for accurately measuring trail segments under real-world conditions. Specifically, we (i) quantify the horizontal and vertical variability introduced by five consumer-grade GNSS devices following a standardized acquisition protocol; (ii) evaluate how sampling resolution and elevation data source affect derived trajectory metrics such as distance, slope, and elevation gain; and (iii) benchmark the resulting pipeline against route profiles generated by widely used online fitness platforms.

II. Methods

A. Data Collection

The field measurement took place on a well-maintained 1 km trail segment in Mt Coot-tha (Brisbane, Australia). Conditions were clear and calm (mean temperature 22 °C, relative humidity 48%, no canopy drip), ensuring stable GNSS reception. Six consecutive laps were completed along the exact same route-already familiar to the lead researcher-to minimize pacing and navigation variability. Figure 1 summarizes the full workflow from acquisition to analysis.

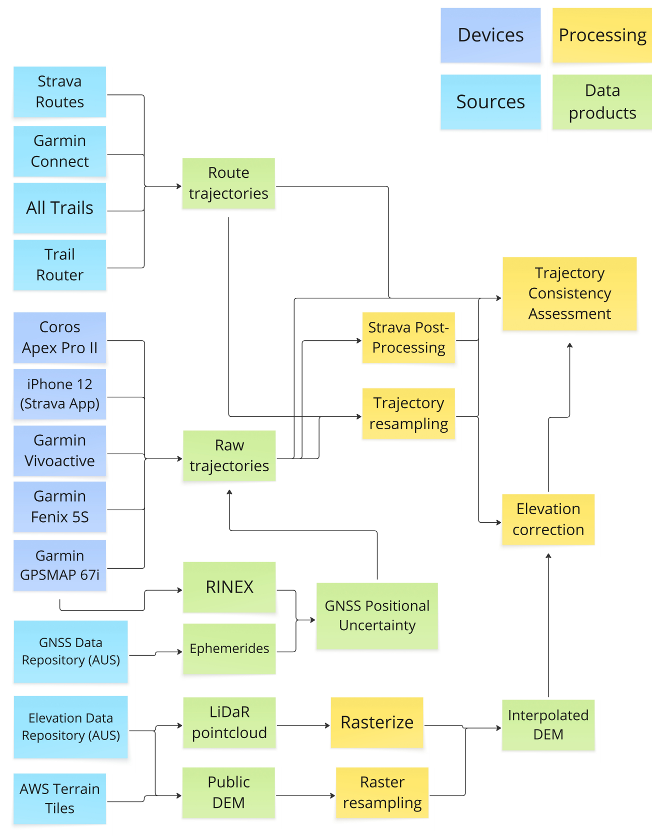

Fig. 1. Methodological workflow. Boxes denote data sources, devices, intermediate products and processing steps; arrows indicate information flow from field acquisition to trajectory assessment.

Devices and Mounting: Five consumer-grade GNSS units were evaluated: *Garmin Fenix 5S*, *Garmin Vivoactive 5*, *Coros Apex Pro II*, *Garmin GPSMAP 67i*, and an *iPhone 12 mini*. All loggers were secured on the top of a backpack, providing unobstructed sky view and eliminating wrist-movement artifacts. Satellite-quality indicators were verified before every lap.

Recording Protocol: Each device stored GNSS positions at 1 Hz (longitude, latitude, ellipsoidal height, timestamp). The *GPSMAP 67i* simultaneously logged dual-frequency raw observations and created RINEX v3.05 files for per-epoch quality auditing. All coordinates were transformed to the WGS 84 / UTM zone 56 S projection prior to metric calculations.

Derived Variables:

- $\Delta\lambda$, $\Delta\varphi$ - longitudinal and latitudinal increments.
- Δh - elevation change.
- Δd_{2D} - surface-distance chord in 2D.
- Δd_{3D} - terrain-following 3D distance.
- Slope - instantaneous gradient ($\Delta h/\Delta d_{2D}$).
- θ_{turn} - heading change between successive segments.

Terrain Data:

- **Public DEMs** - SRTM 1 arc-sec (30 m) tiles from AWS Terrain, plus open-access local DEMs at 5 m and 1 m grid spacing.

- **Lidar point clouds** - Queensland Government surveys from 2009 (2.8 pts/m^2), 2014 (19.3 pts/m^2) and 2019 (46.6 pts/m^2).

All elevation models were re-projected to WGS 84 / UTM 56 S and resampled with bilinear interpolation to match the horizontal resolution required by each analysis step.

GNSS ephemerides: Precise orbit and satellite clock products for each measurement day were downloaded from the Queensland Spatial Catalogue (QSpatial). QSpatial distributes International GNSS Service (IGS) rapid ephemerides in SP3 format-providing satellite positions, velocities and clock corrections at 15 min intervals-as well as 1 Hz satellite clock corrections in CLK files. These products were time-matched to the RINEX observation files recorded by the *GPSMAP 67i* and ingested into the GNSS-processing pipeline to improve satellite-position accuracy and mitigate broadcast-ephemeris errors of up to ~ 2 m in 3D.

B. Device-Induced Variability in GNSS Trajectories

We quantified the horizontal and vertical variability introduced by five consumer-grade GNSS devices following an identical protocol. We first computed reading-level metrics for each pair of consecutive points: elevation change, geodetic (2D) distance, true (3D) distance, and turn angle. Numerical summaries and graphical visualizations were produced for every device and processing stage. Next, the recorded tracks were overlaid to enable visual inspection of spatial discrepancies under identical environmental and operational conditions. Then we assessed how device choice and platform-based processing affect derived trajectory metrics such as distance, slope, and elevation gain.

First we compare: (i) raw GNSS output; (ii) tracks processed by Strava's automatic track-sanitation service; and (iii) tracks whose elevation values were replaced using Strava's proprietary elevation-correction service.

a) Spatial consistency metrics: For each set of six trajectories we computed, in both two and three dimensions:

- **Hausdorff distance** (d_H).

$$d_H(A, B) = \max\Big\{\sup_{a\in A} \inf_{b\in B} \|a - b\|,\ \sup_{b\in B} \inf_{a\in A} \|b - a\|\Big\}.$$

- **Separation** (d_S).

$$d_S(A, B) = \operatorname{median}\Big\{\sup_{a\in A} \inf_{b\in B} \|a - b\|,\ \sup_{b\in B} \inf_{a\in A} \|b - a\|\Big\}.$$

- **Metric Score** (MS). Percentage of points whose nearest-neighbor distance to the opposite trajectory is ≤ 1 m.

For every metric we reported the mean and the coefficient of variation (CV%) across the fifteen pairwise comparisons in each trajectory set.

b) Route-derived attributes: Consistency across rounds was also assessed for:

- total 2D and 3D distance,
- elevation gain and loss,
- maximum and minimum elevation,
- mean slope and inter-quartile range,
- sinuosity (path length divided by straight-line distance),
- mean turn angle.

C. Impact of Sampling Resolution and Elevation Source on Route Metrics

To address Objective 2, we evaluated how sampling resolution and elevation data influence derived trajectory metrics.

a) Spatial resolution resampling: We applied a uniform arc-length parametrization to the raw GNSS data, a resampling technique that redistributes vertices along each trajectory at fixed chord lengths [9]. By decoupling the record from the devices' 1 Hz time-based sampling, this method enforces equal geodesic spacing between successive points.

Trajectories were resampled at nine step sizes: 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4 and 5 m. For every resolution the following attributes were recomputed:

- total 2D and 3D distance
- sinuosity and mean turn angle
- mean slope and slope inter-quartile range
- cumulative elevation gain and loss

Comparing these outputs across scales reveals which metrics are most sensitive to horizontal sampling and quantifies the fractal behavior inherent in trail-running trajectories.

b) DEM Source and Resolution: We then analyzed two families of Digital Elevation Models (DEMs) to quantify the effect of vertical resolution on elevation-driven metrics.

- **Public raster DEMs**: open SRTM dataset resampled bilinearly to grid spacings of 0.25, 0.5, 1, 2, 4, 8, 16, and 32 m.
- **Lidar-derived DEMs**: airborne lidar point clouds from the 2009, 2014, and 2019 Queensland campaigns, rasterized via a TIN interpolation.

All GNSS raw trajectories were first resampled to a uniform 1 m horizontal step to mitigate the fractal complexity distortion. Standardized trajectory elevations were then replaced with the heights extracted from each DEM under test. The following elevation-dependent metrics were recomputed:

- total elevation gain and loss
- maximum and minimum elevation
- mean slope and slope inter-quartile range
- total 3D path length

Comparing these outputs across grid sizes and data sources highlights which DEM configurations yield the most consistent elevation-driven metrics while balancing processing cost and data availability.

D. Benchmarking Best-Practice Pipeline Against Online Route Builders

To address Objective 3, we benchmarked the standardized field-measured tracks against five routes generated by popular online platforms.

For every epoch, east, north and vertical standard deviations were extracted from the RINEX logs and precise ephemerides, then combined into a single 3D uncertainty:

$$\sigma_{3D}(t) = \sqrt{\sigma_E^2(t) + \sigma_N^2(t) + \sigma_V^2(t)},$$

where $\sigma_E(t)$, $\sigma_N(t)$ and $\sigma_V(t)$ denote the east, north and vertical standard deviations at epoch t.

Two *GPSMAP 67i* trajectories with the lowest positional uncertainty served as references, each maintaining $\sigma_{3D} < 0.45$ m across 100% of epochs.

These two reference tracks were compared against all other measurements from the same device, as well as against all routes generated by the platforms. Platform routes were downloaded from AllTrails, Garmin Connect, Trail Router, and Strava (automatic and manual heat-map modes). All datasets were projected to WGS 84 / UTM zone 56 S, resampled at 1 m spacing, and assigned elevation values from the 1 m 2019 lidar DEM.

Spatial fidelity was assessed with:

- total 3D length difference,
- mean absolute separation,
- 3D Hausdorff distance,
- cumulative elevation-gain and -loss error,
- mean-slope bias and inter-quartile range.

These comparisons quantify how closely platform-generated routes replicate the geometry and elevation profile of the best available field data.

III. Results

Figure 2 summarises segment-wise variability for four metrics under three processing modes (Raw, Strava slope sanitiser, Strava elevation correction). The *GPSMAP 67i* shows the tightest dispersions, whereas the *iPhone 12 mini* and *Vivoactive 5* display wider spreads, especially in ΔElevation and turn angle. Integer-rounded heights in several watches introduce ±1 m steps that mask raw slope variability. Strava elevation correction compresses vertical dispersion but leaves planar metrics unchanged, confirming altitude-only adjustment.

Raw trajectory overlays (Figure 3) reinforce these findings: *GPSMAP 67i* and iPhone traces overlap almost perfectly, while *Apex Pro II* and *Vivoactive 5* drift, notably on curves and low-elevation sections.

Fourteen aggregated metrics (Figure 4) show that slope sanitization lowers the CV of elevation gain/loss for all devices; DEM-based correction tightens these metrics further and reduces 3-D separation in four of the five devices, yet leaves horizontal distances virtually unchanged. Mean slope is stable, but its IQR widens after correction due to the finer vertical detail injected from the DEM. Device sensitivity diverges: *GPSMAP 67i* remains largely unaffected, whereas the iPhone experiences the largest relative shifts.

Resampling analysis (Figure 5) reveals that 2-D/3-D lengths stay flat below the native 1 m spacing and decline monotonically at coarser intervals. Elevation gain varies by $\sim 8\%$; slope IQR peaks at 1–2 m, indicating maximal terrain expressiveness at that scale. Sinuosity and turn angle decrease steadily with step size.

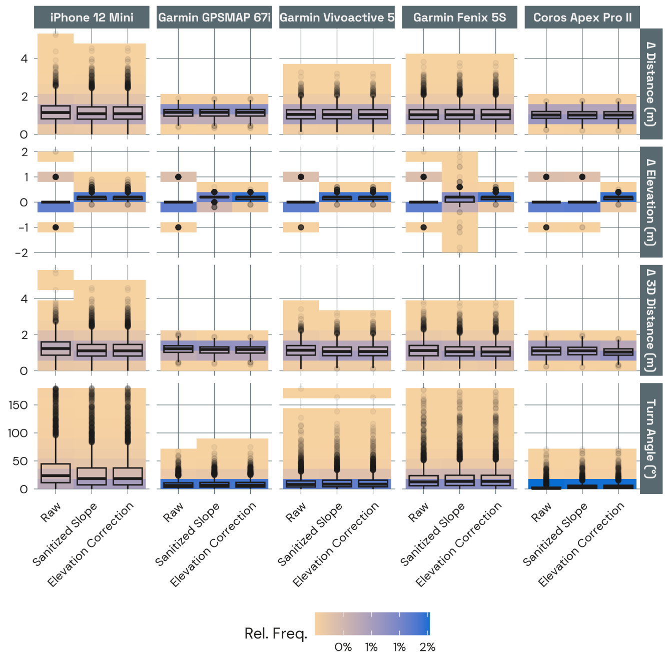

Fig. 2. Reading-wise distributions of four differential metrics by device and processing mode. Heatmaps indicate relative frequency, box plots show quartiles and median, and black dots denote outliers.

DEM resolution effects (Figure 6) are pronounced for public rasters but negligible for lidar surfaces. Resampled public DEMs underestimate elevation gain by up to 12 m and collapse slope variability at coarse grids, while lidar-derived metrics remain stable from 0.25 m to 30 m.

Benchmarking (Figure 7) confirms empirical tracks outperform synthetic routes. Best Rounds achieve 3-D Hausdorff 2.7 m, mean separation 0.6 m and Metric Score 88 %. The best device averaged over six laps yields Hausdorff 5 m and Metric Score 31 %. Route Builder tracks exceed 30 m Hausdorff and fall below 3 % Metric Score, underestimating elevation gain by up to 12 m even after DEM correction. These gaps highlight the limited geometric fidelity of virtual route planners.

IV. Discussion

This study set out to determine how device choice, processing workflow and spatial resolution influence the accuracy of trail-running trajectories. Several studies have documented the spatial inaccuracies of consumer-grade GNSS devices under real-world conditions, with reported errors in total distance and elevation gain ranging from 3-9% [15], [16]. Our results reinforce these findings by demonstrating that even under ideal reception, substantial variation persists across devices and metrics. However, unlike prior studies that relied on single-mode comparisons or limited device types, this work systematically benchmarks five devices across multiple processing pipelines and reference elevations, highlighting the conditions under which sub-meter repeatability is achievable.

Segment-wise analyses (Figure 2) revealed a clear performance hierarchy. The *GPSMAP 67i* achieved sub-meter re-

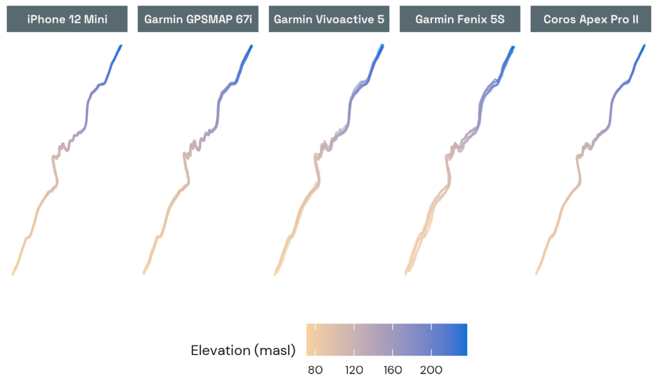

Fig. 3. Raw trajectories captured by five GNSS devices over six repeated laps, colored by absolute elevation. Each panel combines the six laps of a single device. The warm-to-cool gradient represents elevation in meters above sea level, allowing simultaneous visual comparison of horizontal dispersion and vertical profile.

Device	Hausdorff 2D: Raw	Hausdorff 2D: Strava Sanitized	Hausdorff 2D: Strava Elevation	Hausdorff 3D: Raw	Hausdorff 3D: Strava Sanitized	Hausdorff 3D: Strava Elevation
Coros Apex Pro II	8.1 (48.0%)	8.1 (47.9%)	8.1 (47.9%)	8.2 (48.3%)	8.9 (41.2%)	8.3 (46.7%)
Garmin Fenix 5S	16.9 (37.8%)	16.9 (37.8%)	16.9 (37.8%)	17.0 (37.9%)	17.4 (37.4%)	16.9 (37.5%)
Garmin Vivoactive 5	11.7 (28.0%)	11.7 (28.1%)	11.7 (28.1%)	11.9 (28.2%)	11.8 (26.8%)	11.8 (26.8%)
Garmin GPSMAP 67i	5.1 (38.2%)	5.1 (38.2%)	5.1 (38.2%)	5.1 (37.0%)	6.5 (37.7%)	6.6 (41.2%)
iPhone 12 Mini	6.6 (34.7%)	6.6 (34.4%)	6.6 (34.4%)	6.7 (34.5%)	7.2 (31.4%)	7.2 (31.4%)

Device	Separation 2D: Raw	Separation 2D: Strava Sanitized	Separation 2D: Strava Elevation	Separation 3D: Raw	Separation 3D: Strava Sanitized	Separation 3D: Strava Elevation
Coros Apex Pro II	1.5 (28.6%)	1.5 (28.7%)	1.5 (28.7%)	1.6 (27.8%)	3.0 (43.4%)	1.9 (28.0%)
Garmin Fenix 5S	5.5 (53.4%)	5.5 (53.3%)	5.5 (53.3%)	5.6 (52.8%)	7.1 (36.5%)	5.8 (48.9%)
Garmin Vivoactive 5	3.6 (46.1%)	3.6 (46.1%)	3.6 (46.1%)	3.6 (45.4%)	4.0 (37.8%)	4.0 (37.8%)
Garmin GPSMAP 67i	2.6 (61.8%)	2.6 (61.9%)	2.6 (61.9%)	2.6 (61.1%)	4.4 (47.8%)	2.8 (56.3%)
iPhone 12 Mini	1.3 (26.8%)	1.3 (26.2%)	1.3 (26.2%)	1.3 (25.1%)	2.2 (28.2%)	2.2 (28.2%)

Device	Metric Score 2D: Raw	Metric Score 2D: Strava Sanitized	Metric Score 2D: Strava Elevation	Metric Score 3D: Raw	Metric Score 3D: Strava Sanitized	Metric Score 3D: Strava Elevation
Coros Apex Pro II	35.1 (34.6%)	35.0 (34.8%)	35.0 (34.8%)	33.3 (35.1%)	5.5 (118.4%)	24.5 (52.8%)
Garmin Fenix 5S	12.2 (101.4%)	12.2 (101.8%)	12.2 (101.8%)	11.5 (101.9%)	3.0 (160.3%)	5.5 (114.2%)
Garmin Vivoactive 5	15.4 (74.2%)	15.5 (74.2%)	15.5 (74.2%)	14.4 (73.6%)	8.0 (68.0%)	8.0 (68.0%)
Garmin GPSMAP 67i	31.6 (110.3%)	31.5 (110.5%)	31.5 (110.5%)	30.1 (111.4%)	5.3 (256.6%)	24.7 (115.1%)
iPhone 12 Mini	40.6 (29.2%)	40.0 (29.0%)	40.0 (29.0%)	37.9 (29.6%)	16.3 (60.6%)	16.3 (60.6%)

Device	Total Length 2D: Raw	Total Length 2D: Strava Sanitized	Total Length 2D: Strava Elevation	Total Length 3D: Raw	Total Length 3D: Strava Sanitized	Total Length 3D: Strava Elevation
Coros Apex Pro II	984.3 (2.7%)	985.2 (2.7%)	985.2 (2.7%)	1057.6 (2.7%)	1053.9 (2.9%)	1002.5 (2.6%)
Garmin Fenix 5S	1026.4 (1.9%)	1027.5 (1.8%)	1027.5 (1.8%)	1095.8 (2.3%)	1056.0 (1.9%)	1043.9 (1.8%)
Garmin Vivoactive 5	1022.5 (1.8%)	1023.5 (1.8%)	1023.5 (1.8%)	1091.5 (2.1%)	1040.0 (1.7%)	1040.0 (1.7%)
Garmin GPSMAP 67i	1081.0 (0.7%)	1081.7 (0.7%)	1081.7 (0.7%)	1151.0 (0.6%)	1100.6 (0.7%)	1097.6 (0.7%)
iPhone 12 Mini	1134.4 (2.1%)	1108.8 (1.5%)	1108.8 (1.5%)	1200.9 (2.1%)	1123.8 (1.4%)	1123.8 (1.4%)

Device	Elevation Gain: Raw	Elevation Gain: Strava Sanitized	Elevation Gain: Strava Elevation	Elevation Loss: Raw	Elevation Loss: Strava Sanitized	Elevation Loss: Strava Elevation
Coros Apex Pro II	170.2 (1.6%)	162.0 (0.8%)	163.0 (0.9%)	-5.8 (-56.8%)	-0.5 (-109.5%)	-0.1 (-0.0%)
Garmin Fenix 5S	170.0 (2.6%)	166.0 (0.7%)	162.1 (1.2%)	-6.7 (-55.0%)	-4.2 (-50.2%)	-0.1 (-0.0%)
Garmin Vivoactive 5	168.5 (1.1%)	161.9 (1.8%)	161.9 (1.8%)	-4.5 (-36.5%)	-0.1 (-0.0%)	-0.1 (-0.0%)
Garmin GPSMAP 67i	173.2 (2.6%)	163.4 (0.5%)	162.3 (1.9%)	-7.2 (-63.3%)	-0.4 (-18.8%)	-0.1 (-0.0%)
iPhone 12 Mini	175.5 (2.5%)	161.2 (1.3%)	161.2 (1.3%)	-10.5 (-38.9%)	-0.1 (-0.0%)	-0.1 (-0.0%)

Device	Max Elevation: Raw	Max Elevation: Strava Sanitized	Max Elevation: Strava Elevation	Min Elevation: Raw	Min Elevation: Strava Sanitized	Min Elevation: Strava Elevation
Coros Apex Pro II	236.5 (0.2%)	248.2 (0.9%)	233.5 (0.6%)	72.2 (1.8%)	86.3 (2.2%)	70.6 (0.1%)
Garmin Fenix 5S	236.3 (0.3%)	250.6 (0.5%)	232.8 (0.8%)	73.0 (2.9%)	88.4 (2.8%)	70.8 (0.4%)
Garmin Vivoactive 5	236.5 (0.2%)	232.7 (1.1%)	232.7 (1.1%)	72.5 (1.9%)	70.8 (0.5%)	70.8 (0.5%)
Garmin GPSMAP 67i	237.0 (0.0%)	234.9 (0.8%)	232.9 (1.3%)	71.0 (0.0%)	71.8 (2.7%)	70.7 (0.5%)
iPhone 12 Mini	236.7 (0.2%)	232.1 (1.0%)	232.1 (1.0%)	71.7 (1.1%)	70.9 (0.9%)	70.9 (0.9%)

Device	Mean Slope: Raw	Mean Slope: Strava Sanitized	Mean Slope: Strava Elevation	IQR Slope: Raw	IQR Slope: Strava Sanitized	IQR Slope: Strava Elevation
Coros Apex Pro II	17.7 (0.9%)	17.8 (1.5%)	17.5 (1.8%)	0.0 (NaN%)	0.0 (NaN%)	15.0 (2.6%)
Garmin Fenix 5S	16.5 (1.6%)	19.5 (1.7%)	16.6 (2.3%)	0.0 (NaN%)	27.5 (3.3%)	13.0 (4.7%)
Garmin Vivoactive 5	16.9 (1.8%)	16.7 (1.7%)	16.7 (1.7%)	0.0 (NaN%)	13.8 (4.6%)	13.8 (4.6%)
Garmin GPSMAP 67i	16.2 (2.4%)	16.4 (4.4%)	16.1 (4.8%)	0.0 (NaN%)	13.2 (58.2%)	14.6 (9.2%)
iPhone 12 Mini	14.4 (4.5%)	15.2 (2.2%)	15.2 (2.2%)	0.0 (NaN%)	11.9 (8.1%)	11.9 (8.1%)

Device	Sinuosity: Raw	Sinuosity: Strava Sanitized	Sinuosity: Strava Elevation	Mean Turn Angle: Raw	Mean Turn Angle: Strava Sanitized	Mean Turn Angle: Strava Elevation
Coros Apex Pro II	1.1 (2.6%)	1.1 (2.6%)	1.1 (2.6%)	2.9 (4.0%)	4.9 (5.5%)	4.9 (5.5%)
Garmin Fenix 5S	1.2 (1.1%)	1.2 (1.1%)	1.2 (1.1%)	18.0 (5.8%)	18.5 (5.2%)	18.5 (5.2%)
Garmin Vivoactive 5	1.2 (1.4%)	1.2 (1.4%)	1.2 (1.4%)	11.0 (13.9%)	11.6 (13.3%)	11.6 (13.3%)
Garmin GPSMAP 67i	1.2 (0.7%)	1.2 (0.7%)	1.2 (0.7%)	8.0 (10.2%)	8.6 (9.6%)	8.6 (9.6%)
iPhone 12 Mini	1.3 (2.3%)	1.3 (1.7%)	1.3 (1.7%)	34.7 (11.6%)	28.4 (9.2%)	28.4 (9.2%)

Fig. 4. Mean values and coefficients of variation (CV%) in parentheses for 14 trajectory metrics, computed for five GNSS devices under three processing modes. Cell shading expresses each CV as a percentage of the full range observed for that metric; lighter blues denote higher relative mean.

peatability (Hausdorff 2.7 m; separation 0.6 m), whereas wrist- and phone-based loggers displayed wider spreads, particularly in turn angle and ΔElevation. Integer-rounded heights in some watches produced artificial zero-slope segments, inflating vertical noise. These findings confirm an *intrinsic quality gap* among consumer units that no post-processing can fully erase.

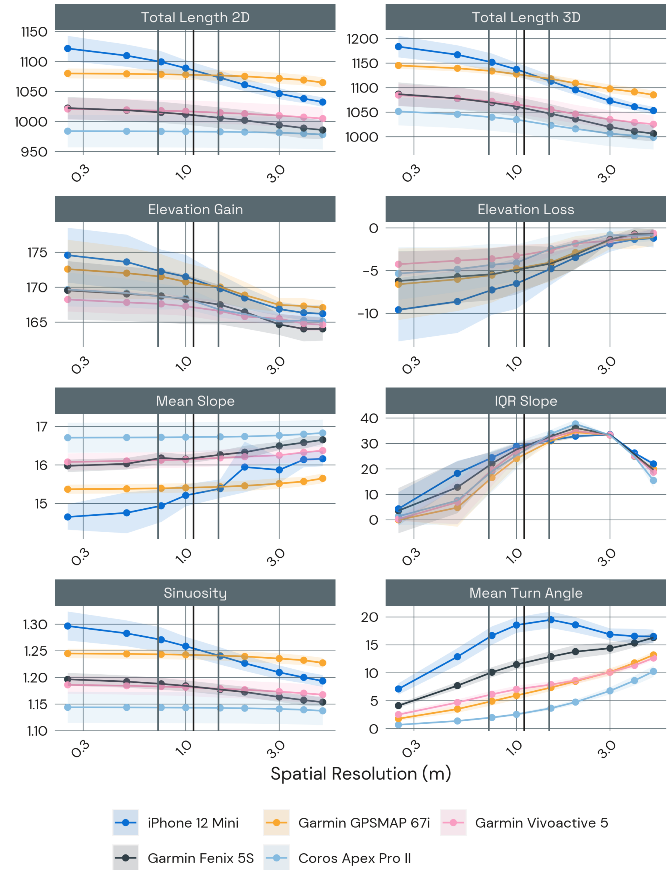

Fig. 5. Sensitivity of derived trajectory metrics to spatial resampling. Lines show the mean per device at each fixed interval; vertical bars mark mean ± standard deviation of native GNSS spatial resolution.

Below the native spacing of 1.1 ± 0.4 m, additional densification yields negligible new information. Above 2.5 m, total length, sinuosity and turn angle diverge markedly across devices (Fig. 5). Slope descriptors never fully stabilize, indicating that fractal effects persist until complemented by a high-resolution DEM.

Lidar-derived DEMs produced nearly flat response curves across all resolutions (Fig. 6). Public rasters, when merely resampled, underestimated elevation gain by up to 12 m and collapsed slope IQR beyond 5 m. A smaller grid suffices to stabilize slope IQR; total 3D length converges by 1 m. *You can't beat good data*: deriving surfaces directly from point clouds is preferable to resampling coarse rasters.

Route-builder tracks exhibited Hausdorff distances >30 m and Metric Scores <3 %, far behind the best rounds (Fig. 7). A single well captured field measurement outperformed both the best-device average and all platform outputs.

Next steps include: (1) extending the protocol to longer, multi-biome trails; (2) testing vendor-provided raw GNSS (RINEX) once smartwatch manufacturers expose it; and (3) integrating real-time kinematic (RTK) corrections to examine whether centimeter-level accuracy translates into meaningful biomechanical insights.

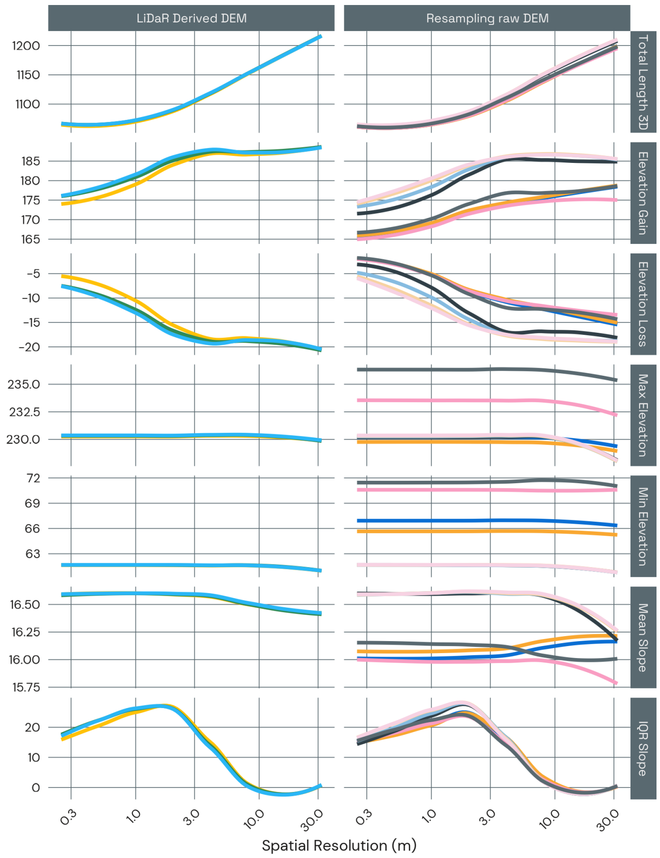

Fig. 6. Effect of DEM source and grid spacing on elevation metrics. Left: Lidar-derived surfaces; right: public rasters resampled to the same spacings (0.25-30 m, log-scale). Lines represent individual DEMs

Metric	Group	Raw	1m-std	1m-std + 1m-DEM
Hausdorff 2D	Best Rounds	2.7 (0.2%)	2.7 (0.1%)	2.7 (0.1%)
	Best Device	5.1 (38.2%)	5.0 (38.4%)	5.0 (38.4%)
	Route Builders	36.0 (64.8%)	31.9 (83.7%)	31.9 (83.7%)
Hausdorff 3D	Best Rounds	2.8 (2.6%)	2.7 (0.1%)	2.7 (0.1%)
	Best Device	5.1 (37.0%)	5.1 (37.5%)	5.1 (37.5%)
	Route Builders	41.4 (46.9%)	39.1 (54.7%)	39.1 (54.7%)
Separation 2D	Best Rounds	0.5 (0.2%)	0.5 (0.0%)	0.5 (0.0%)
	Best Device	2.6 (61.8%)	2.6 (62.5%)	2.6 (62.5%)
	Route Builders	3.5 (82.0%)	1.1 (42.6%)	1.1 (42.6%)
Separation 3D	Best Rounds	0.5 (0.7%)	0.6 (0.1%)	0.6 (0.1%)
	Best Device	2.6 (61.1%)	2.6 (61.3%)	2.6 (61.3%)
	Route Builders	8.1 (38.7%)	6.4 (56.3%)	6.4 (56.3%)
Metric Score 2D	Best Rounds	88.9 (0.2%)	89.7 (0.0%)	89.7 (0.0%)
	Best Device	31.6 (110.3%)	32.1 (109.0%)	32.1 (109.0%)
	Route Builders	26.6 (100.7%)	53.4 (41.7%)	53.4 (41.7%)
Metric Score 3D	Best Rounds	86.1 (1.0%)	88.2 (0.2%)	88.2 (0.2%)
	Best Device	30.1 (111.4%)	31.0 (109.8%)	31.0 (109.8%)
	Route Builders	2.7 (114.8%)	5.6 (70.6%)	5.6 (70.6%)
Total Length 2D	Best Rounds	1071.4 (0.2%)	1068.0 (0.2%)	1068.0 (0.2%)
	Best Device	1081.0 (0.7%)	1078.2 (0.8%)	1078.2 (0.8%)
	Route Builders	1008.4 (2.6%)	972.3 (3.6%)	972.3 (3.6%)
Total Length 3D	Best Rounds	1147.9 (0.6%)	1120.6 (0.3%)	1092.7 (0.2%)
	Best Device	1151.0 (0.6%)	1127.6 (0.6%)	1103.2 (0.8%)
	Route Builders	1028.1 (2.6%)	991.7 (3.7%)	998.6 (3.6%)
Elevation Gain	Best Rounds	174.5 (2.8%)	171.2 (1.4%)	178.5 (0.4%)
	Best Device	173.2 (2.6%)	170.8 (1.6%)	180.0 (1.0%)
	Route Builders	155.4 (5.5%)	153.2 (5.6%)	176.8 (0.6%)
Elevation Loss	Best Rounds	-8.5 (-58.2%)	-5.2 (-44.9%)	-9.7 (-7.7%)
	Best Device	-7.2 (-63.3%)	-4.8 (-56.6%)	-11.1 (-15.7%)
	Route Builders	-3.7 (-186.0%)	-3.7 (-186.9%)	-11.8 (-16.7%)
Max Elevation	Best Rounds	237.0 (0.0%)	237.0 (0.0%)	230.5 (0.0%)
	Best Device	237.0 (0.0%)	237.0 (0.0%)	230.4 (0.0%)
	Route Builders	225.8 (4.5%)	225.8 (4.5%)	230.9 (0.0%)
Min Elevation	Best Rounds	71.0 (0.0%)	71.0 (0.0%)	61.6 (0.1%)
	Best Device	71.0 (0.0%)	71.0 (0.0%)	61.5 (0.3%)
	Route Builders	74.8 (6.6%)	74.8 (6.6%)	63.6 (2.9%)
Mean Slope	Best Rounds	16.6 (0.7%)	15.6 (0.2%)	15.8 (0.2%)
	Best Device	16.2 (2.4%)	15.4 (0.8%)	15.7 (0.8%)
	Route Builders	73.2 (174.3%)	15.5 (8.6%)	17.0 (3.8%)
IQR Slope	Best Rounds	0.0 (NaN%)	21.4 (9.7%)	20.5 (3.5%)
	Best Device	0.0 (NaN%)	24.1 (12.5%)	20.8 (4.8%)
	Route Builders	16.6 (14.7%)	17.6 (16.0%)	20.2 (4.1%)
Sinuosity	Best Rounds	1.2 (0.3%)	1.2 (0.3%)	1.2 (0.3%)
	Best Device	1.2 (0.7%)	1.2 (0.7%)	1.2 (0.7%)
	Route Builders	1.2 (4.7%)	1.2 (2.8%)	1.2 (2.8%)
Mean Turn Angle	Best Rounds	8.7 (1.0%)	6.6 (2.3%)	6.6 (2.3%)
	Best Device	8.0 (10.2%)	5.9 (12.6%)	5.9 (12.6%)
	Route Builders	34.5 (30.1%)	1.5 (20.6%)	1.5 (20.6%)

Fig. 7. Metric comparison for: Best Rounds, Best Device (GPSMAP 67i, all laps), and Route Builders (five online platforms). Cells show mean value with CV % in parentheses; blue shading indicates relative deviation within the full observed range. Metrics are reported for raw tracks, 1 m-sanitized tracks, and 1 m-sanitized + lidar-DEM tracks.

By combining open RINEX logging, raw DEM evaluation and resolution-sensitivity testing, this study offers a fully transparent workflow that can be replicated, audited, and extended. This contributes to ongoing efforts to build validated geospatial baselines for locomotion research, where terrain metrics like true 3D distance or slope variability are increasingly linked to physiological modeling and wearable data integration.

To our knowledge, this is the first openly documented workflow that combines (i) dual-frequency surveying, (ii) resolution-sweep analysis and (iii) DEM source comparison within one reproducible pipeline. The benchmark results establish practical targets: Hausdorff <3 m, separation <1 m, for future locomotion studies.

The results of this study must be interpreted considering certain limitations. Data collection was limited to a single researcher performing six repeated laps on a 1 km trail under ideal GNSS conditions, prioritizing control and repeatability over generalization. No statistical inference or sample size estimation was conducted; results are based on descriptive metrics and spatial comparisons. The evaluated devices reflect common consumer-grade hardware, but the roster was not exhaustive and did not include professional survey receivers (e.g., dual-frequency RTK/PPK rovers or total-station-grade instruments) that could set an even tighter ground-truth benchmark. The test course was a single 1 km loop segment; results may differ on longer routes, in deep canyons, or in dense forest where satellite geometry is poorer. As such, findings should be interpreted as a benchmark of relative device and processing performance under controlled conditions.

A. Conclusions

High-quality data, an appropriate device and a transparent processing chain together enable repeatable, sub-meter trail measurements. However, achieving full ecological validity in human movement studies will require combining detailed terrain models with complementary technologies that capture physiological and biomechanical signals (e.g., heart rate, oxygen consumption, gait cycle), as well as exploring analytical frameworks that leverage wearable sensors and predictive modeling [27]. Beyond biomechanics, precise terrain mapping holds value for mountain sport science, military applications, environmental risk assessment, and public health initiatives that promote trail-based physical activity. A transparent geospatial baseline allows these fields to make consistent comparisons and informed decisions.

References

[1] G. Vernillo, G. P. Millet, and J. Giandolini, "Biomechanics and physiology of uphill and downhill running," *Sports Med.*, vol. 47, no. 4, pp. 615–629, 2017, doi: 10.1007/s40279-016-0605-y.

[2] J. Saugy, G. Place, S. Millet, et al., "Alterations of neuromuscular function after the world's most challenging mountain ultra-marathon," *PLoS One*, vol. 8, no. 6, p. e65596, 2013, doi: 10.1371/journal.pone.0065596.

[3] G. Björklund, M. Swarén, D.-P. Born, and T. Stöggl, "Biomechanical adaptations and performance indicators in short trail running," *Front. Physiol.*, vol. 10, p. 506, 2019, doi: 10.3389/fphys.2019.00506.

[4] R. Sánchez, "Analysing gait transition dynamics in uphill trail running from wearable devices," *ISBS Proc. Arch.*, vol. 42, no. 1, p. 803, 2024.

[5] R. Sánchez and C. Nieto-Jiménez, "Use of digital devices to follow a trail runner: case study," *Retos: Nuevas Perspect. Educ. Fís. Deporte Recreac.*, no. 83, pp. 582–586, 2020, doi: 10.47197/retos.v0i83.82633.

[6] L. H. Palucci Vieira, C. Carling, and P. S. Bradley, "Should GPS data be normalized for performance and fatigue monitoring in soccer? A theoretical-practical discussion on high-speed running," *Front. Sports Act. Living*, vol. 7, p. 1603767, 2025, doi: 10.3389/fspor.2025.1603767.

[7] P. J. Bascuas, C. Comeras-Chueca, C. Berzosa, H. Gutiérrez, J. Del Coso, and A. Álvarez-Mesa, "Modifications of viscoelastic properties and physiological parameters after performing uphill and downhill running trials," *J. Sports Med. Phys. Fitness*, vol. 61, no. 4, pp. 625–635, 2021, doi: 10.23736/S0022-4707.20.11089-6.

[8] C. Berzosa, C. Comeras-Chueca, P. J. Bascuas, H. Gutiérrez, and A. V. Bataller-Cervero, "Assessing trail running biomechanics: a comparative analysis of the reliability of Stryd™ and GARMIN^{RP} wearable devices," *Sensors*, vol. 24, no. 11, p. 3570, 2024, doi: 10.3390/s24113570.

[9] R. Sánchez, P. E. Egli, K. Jornet, M. Duggan, and M. Besomi, "How fractal complexity distorts distance and elevation gain in trail and mountain running: the case for course measurement standardization," *Proc. Inst. Mech. Eng. P J. Sports Eng. Technol.*, 2025, doi: 10.1177/17543371251341660.

[10] J. Huang, J. Zhang, L. Li, and Y. Wang, "Evaluation of positioning accuracy of smartphones under different canopy openness," *Forests*, vol. 13, no. 10, p. 1591, 2022, doi: 10.3390/f13101591.

[11] B. Wessel, M. Huber, C. Wohlfart, U. Marschalk, D. Kosmann, and A. Roth, "Accuracy assessment of the global TanDEM-X digital elevation model with GPS data," *ISPRS J. Photogramm. Remote Sens.*, vol. 139, pp. 171–182, 2018, doi: 10.1016/j.isprsjprs.2018.02.017.

[12] R. Sánchez, P. Egli, M. Besomi, and R. Truffello, "Assessing the impact of digital elevation model resolution on elevation gain estimations in trail running," in *Proc. 2024 Int. Conf. Electr., Comput. Energy Technol. (ICECET)*, 2024, doi: 10.1109/ICECET59688.2024.10423456.

[13] S. J. Jalal, A. A. Abdulhassan, and S. A. Al-Saadi, "Optimizing the global digital elevation models (GDEMs) and accuracy of derived DEMs from GPS points for Iraq's mountainous areas," *Geod. Geodyn.*, vol. 11, no. 5, pp. 338–349, 2020, doi: 10.1016/j.geog.2020.07.005.

[14] X. Liu, M. Ran, H. Xia, and M. Deng, "Evaluating vertical accuracies of open-source digital elevation models over multiple sites in China using GPS control points," *Remote Sens.*, vol. 14, no. 9, p. 2000, 2022, doi: 10.3390/rs14092000.

[15] R. Gilgen-Ammann, T. Schweizer, and T. Wyss, "Accuracy of distance recordings in eight positioning-enabled sport watches: instrument validation study," *JMIR mHealth uHealth*, vol. 8, no. 6, p. e17118, 2020, doi: 10.2196/17118.

[16] R. Sánchez and M. Villena, "Comparative evaluation of wearable devices for measuring elevation gain in mountain physical activities," *Proc. Inst. Mech. Eng. P J. Sports Eng. Technol.*, vol. 234, no. 6, pp. 522–530, 2020, doi: 10.1177/1754337120922546.

[17] J. L. Carrasco, "Assessing repeatability of spatial trajectories," *Methods Ecol. Evol.*, vol. 15, no. 1, pp. 144–152, 2024, doi: 10.1111/2041-210X.14266.

[18] M. J. Campbell, S. L. Cutler, and P. E. Dennison, "A singular, broadly-applicable model for estimating on- and off-path walking travel rates using airborne lidar data," *Sci. Rep.*, vol. 14, p. 21393, 2024, doi: 10.1038/s41598-024-71359-6.

[19] F. Espinoza and R. Owens, "Identifying roads and trails hidden under canopy using LiDAR," in *Proc. ASPRS 2007 Annu. Conf.*, 2007.

[20] M. Abdelazeem, M. Elshewy, and A. Elsheshtawy, "Towards an accurate real-time digital elevation model using various GNSS techniques," *Sensors*, vol. 24, no. 24, p. 8147, 2024, doi: 10.3390/s24248147.

[21] K. J. Wilson and N. Wilson, "Assessing course difficulty and the effect of weather in amateur cross country running races," *arXiv preprint arXiv:2405.09865*, 2024.

[22] T. Emig and J. Peltonen, "Human running performance from real-world big data," *Nat. Commun.*, vol. 11, p. 4936, 2020, doi: 10.1038/s41467-020-18737-6.

[23] A. D. Townshend, C. J. Worringham, and I. B. Stewart, "Spontaneous pacing during overground hill running," *Med. Sci. Sports Exerc.*, vol. 42, no. 1, pp. 160–169, 2010, doi: 10.1249/MSS.0b013e3181af21e2.

[24] C. S. Easthope, K. Nosaka, C. Caillaud, F. Vercruyssen, J. Louis, and J. Brisswalter, "Reproducibility of performance and fatigue in trail running," *J. Sci. Med. Sport*, vol. 17, no. 2, pp. 207–211, 2014, doi: 10.1016/j.jsams.2013.03.009.

[25] C. Cummins, R. Orr, H. O'Connor, and C. West, "Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review," *Sports Med.*, vol. 43, no. 10, pp. 1025–1042, 2013, doi: 10.1007/s40279-013-0069-2.

[26] T. U. Macfarlane, T. J. Scott, and V. G. Kelly, "The validity and reliability of global positioning systems in team sport: a brief review," *J. Strength Cond. Res.*, vol. 30, no. 5, pp. 1470–1490, 2016, doi: 10.1519/JSC.0000000000001221.

[27] L. Xiang, A. Wang, Y. Gu, L. Zhao, V. Shim, and J. Fernandez, "Recent machine learning progress in lower limb running biomechanics with wearable technology: a systematic review," *Front. Neurorobot.*, vol. 16, p. 913052, 2022, doi: 10.3389/fnbot.2022.913052.