Reliability of running gait variability measures calculated from inertial measurement units

JONES, Ben DM, WHEAT, Jonathan <http://orcid.org/0000-0002-1107-6452>, MIDDLETON, Kane, CAREY, David L and HELLER, Ben <http://orcid.org/0000-0003-0805-8170>

Title

Reliability of running gait variability measures calculated from inertial measurement units.

Authors

Ben Jones:
Sport and Physical Activity Research Centre, Sheffield Hallam University, Olympic Legacy Park, 2 Old Hall Rd, Sheffield, S9 3TY, United Kingdom.
Sport, Performance, and Nutrition Research Group, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, VIC 3086, Australia.
b.d.jones@shu.ac.uk
MSc, BSc

Jon Wheat:
Sport and Human Performance Enhancement Research Centre, Nottingham Trent University, Clifton Campus, Nottingham, NG11 8NS, United Kingdom.
jon.wheat@ntu.edu.au
PhD, BSc (Hons)

Kane Middleton
Sport, Performance, and Nutrition Research Group, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, VIC 3086, Australia.
k.middleton@latrobe.edu.au
PhD, BSc(Hons), BSc

David Carey
Sport, Performance, and Nutrition Research Group, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, VIC 3086, Australia.
d.carey@latrobe.edu.au
PhD, MSc, BSc

Ben Heller*
Sport and Physical Activity Research Centre, Sheffield Hallam University, Olympic Legacy Park, 2 Old Hall Rd, Sheffield, S9 3TY, United Kingdom.
b.heller@shu.ac.uk
PhD, MA (Cantab), BA

* Corresponding author: Ben Heller.

1. Introduction

Running is a popular form of physical activity and provides substantial health benefits (Dai et al., 2015; Lee et al., 2017; Tudor-Locke et al., 2010). However, running entails a high injury risk compared to other sports (Kemler et al., 2018), with running-related injury incidence between 2.5 and 33 per 1000 hours of running (Videbæk et al., 2015). Running-related injuries can result in temporary or permanent cessation of running, and financial costs due to medical treatment or absence from work (Worp et al., 2015). Wearable devices ("wearables"), such as inertial measurement units (IMUs), enable the long-term monitoring of running biomechanics outdoors, and may support prevention and rehabilitation of running-related injuries (Willy, 2018).

Running and running-related injury arise from the non-linear interaction of independent biological systems, and the interaction of the runner with their task and environment (Davids et al., 2003; Fonseca et al., 2020). A high-order signal that emerges from these non-linear interactions may describe the macroscopic dynamic behaviour of the individual and be sensitive to changes in system dynamics that might precede or result from running-related injury (Fonseca et al., 2020). Indeed, a high-order signal that emerges from a healthy human system will contain fluctuations since the system will utilise its abundant degrees of freedom (Stergiou and Decker, 2011). Degrees of freedom define the number of potential configurations of the components at each level of a system (Newell and Vaillancourt, 2001). Stride time (Brahms et al., 2022; Mann et al., 2015; Meardon et al., 2011) and centre-of-mass or sacral acceleration (Gruber et al., 2021; Schütte et al., 2015) are two common choices of high-order variable in running gait literature; they are a function of the multiple interacting subsystems required to run and are relatively easy to measure compared to other possible high-order variables, such as leg stiffness.

Linear measures of variability such as standard deviation (SD) or coefficient of variation (CV) have traditionally been used to quantify the magnitude of fluctuations in high-order signals. However, these metrics do not capture the temporal dynamics of running as they consider strides to be

independent events and thus ignore the role of perception-action coupling in locomotion (Davids et al., 2003). Detrended fluctuation analysis (DFA) (Peng et al., 1995) and sample entropy (Richman and Moorman, 2000) are two non-linear measures of variability that capture these temporal dynamics.

DFA quantifies the complexity of a high-order signal by evaluating its fractal structure and detecting statistical persistence. The DFA algorithm yields the exponent α, with $0 < \alpha < 0.5$ indicating anti-persistence, and $0.5 < \alpha < 1.0$ indicating persistence. Special cases occur at $\alpha = 0.5$, for which a signal consists of white noise and values are completely uncorrelated, and at $\alpha = 1.0$, for which a signal exhibits 1/f behaviour (pink noise).

Stride time time series can be analysed using DFA since they satisfy its requirement for non-cyclical data (Wilson and Likens, 2023). Researchers have found stride time DFA-α to be lower in runners with history of injury (Meardon et al., 2011), to decrease with fatigue (Brahms et al., 2022) and to decrease following a period of heavy training (Bellenger et al., 2019). In these studies, differences of the magnitude 0.11 to 0.17 were found between groups (Meardon et al., 2011) or within individuals (Bellenger et al., 2019; Brahms et al., 2022). However, research into the between-day reliability of running stride time DFA has been limited (Fuller et al., 2018; Godin et al., 2024; Mo and Chow, 2019). An understanding of this is vital for longitudinal monitoring since any between-day changes must be known to be real.

Mo and Chow (2019) found that stride time DFA-α demonstrated good between-day reliability (ICC = 0.81) within an exhaustive thirty-minute run at a constant speed, but poor between-day reliability (ICC: 0.20 – 0.37) when that run was segmented into three sections. In contrast, Godin et al. (2024) and Fuller et al. (2018) found that DFA-α demonstrated moderate to good reliability (ICC: 0.71 – 0.87) during six minute trials at three fixed speeds. However, Fuller et al. (2018) conducted their analyses on time series of 400 strides, which is below the 500 to 600 stride recommendation to achieve reliable estimates of DFA-α (Ravi et al., 2020), and both Fuller et al.

(2018) and Mo and Chow (2019) conducted their studies on 10 or fewer participants. During walking, between-day reliability of stride time DFA-α has been poor (Ryan et al., 2020), moderate (Raffalt et al., 2018) and good (Di Bacco and Gage, 2023), although the applicability of these results to running is uncertain given that individuals display less persistence in walking (Wilson et al., 2023). It is thus evident that further research is needed to assess the reliability of DFA-α when applied to running stride times, with the use of a larger sample size and across a greater range of running conditions than those of Fuller et al. (2018) and Mo and Chow (2019).

Like DFA, the between-day reliability of the sample entropy of biomechanical signals during running is also unknown, despite it being the focus of a proof-of-concept study on monitoring gait variability (Gruber et al., 2021). Sample entropy evaluates a signal's regularity by calculating the conditional probability that two similar vectors of *m* data points will remain similar with the addition of a consecutive data point. Data points are defined as similar if the magnitude of their difference is less than a threshold, set as a percentage, *r,* of the signal's SD. A lower entropy value reflects greater regularity. Sample entropy values from trunk acceleration signals have been found to decrease with fatigue in runners with a history of medial tibial stress syndrome but not in healthy controls (Schütte et al., 2018), and to be lower in lower limb joint angle time histories from runners with running-related injury history compared to those without (Quirino et al., 2021). Nevertheless, to the best of our knowledge, the between-day reliability of sample entropy when applied to a high-order sacral acceleration signals from running gait has not been investigated.

Therefore, the purpose of this study was to investigate the between-day reliability of variability measures when applied to two high-order signals measured by a wearable device: stride times, using CV and DFA, and sacral accelerations, using sample entropy. A range of speed and gradient combinations were investigated to assess the between-day reliability of variability measures across running conditions representative of different overground running scenarios.

2. Methods

2.1 Participants and set-up

28 recreational runners (17 male, 11 female; mean ± SD, age 40.8 ± 14.1 years, mass 72.2 ± 10.1 kg, height 1.77 ± 0.08 m) provided written informed consent to participate in a protocol that received approval from Sheffield Hallam University Ethics Committee and reciprocal approval from La Trobe University Human Ethics Committee. Participants were required to be running a minimum of twice per week, to be able to run 10 km, and to have experienced no running-related musculoskeletal injuries within the previous three months. A RunScribe™ (Scribe Labs, Moss Beach, CA, USA) IMU sampling at 500 Hz and programmed by RunScribe™ (RunScribe™ Blue Plus, 37.53) with propriety algorithms was placed in a lace cradle on each participant's left and right running shoes. These IMUs output processed stride time data at a stride-by-stride level. Stride times from RunScribe™ propriety algorithms have previously been shown to demonstrate excellent validity (Koldenhoven and Hertel, 2018) and reliability (Kozinc et al., 2022). A third RunScribe™ IMU (RunScribe™ Red) collecting and outputting raw acceleration and angular velocity data at 200 Hz was placed in a cradle secured onto the waistband over the sacrum.

2.2 Protocol

Participants completed an eight-minute warm-up at a self-selected running speed, which also served as familiarisation to treadmill running (Van Hooren et al., 2020) (TRM 700, Precor, Woodinville, WA, USA). Participants' preferred running speeds were found using the method described by Jordan et al. (2006). Participants then completed an ordered set of six trials at speeds relative to a self-reported 5 km speed (Table 1). The length of each trial ensured 550 strides were recorded, excluding a minimum of 15 seconds at the beginning and end. This met guidelines on minimum time series length to calculate both stride time DFA-α (Ravi et al., 2020) and sacral acceleration sample entropy (Yentes and Raffalt, 2021). The order of trials was fixed to minimise the impact of order effects on fatigue, which may influence running gait complexity

(Meardon et al., 2011; Mo and Chow, 2018). Participants were given at least 2 minutes and up to 10 minutes recovery between trials (Jordan et al., 2007). Participants returned 7 days later (n = 1, 8 days) to repeat the ordered set of trials.

2.3 Data Processing

Stride time data from the right foot and raw sacral accelerometer data were downloaded from the RunScribe™ dashboard and imported into MATLAB (R2023b, Mathworks Inc, MA, USA).

2.3.1 Stride Times

Trials were expected to contain a minimum of 550 consecutive valid strides. However, every trial had at least one instance of a stride being skipped or erroneous data being reported. Skipped and erroneous strides were uniformly distributed throughout trials. Therefore, the first 512 strides in each trial for which valid data were available were concatenated to form a series of 512 stride times. This concatenation was expected to have a negligible effect on the calculation of DFA-α since stride times in healthy runners display statistical persistence (Wilson and Likens, 2023) and such series are insensitive to concatenation in the cases of data loss (Pavlova et al., 2019). Stride times were subsequently removed if they corresponded to step frequencies lower than 130 steps per minute (spm) or greater than 210 spm. These thresholds were set respectively 10 spm lower and higher than reported cadence thresholds for running (Hansen et al., 2018; Ueno et al., 2021) to account for the variability in stride times which may not be reflected in one-minute averages. A series of 512 stride times could be constructed for all 12 trials (6 trials * 2 days) for 19 participants. Series could not be constructed due to injury between the two days (n = 2), incompletion of trials (n = 3), and data recording errors (n = 4).

The mean and coefficient of variation (CV) of stride times were then calculated. Stride time DFA-α was estimated using the evenly spaced averaged algorithm (Liddy and Haddad, 2018). Parameters were selected to minimise the root mean square error between theoretical and estimated α values for 100 time series of simulated (Kroese and Botev, 2013) fractional Brownian motion and fractional Gaussian noise of length N = 512. α values ranged from 0.3 to 1.4 in steps

of 0.1, excluding 1.0 (Delignières, 2015). All other parameter values included in this testing procedure were those presented by Phinyomark et al. (2020). Consequently, window sizes selected ranged from 4 to N/5 and detrending was performed using a first-order polynomial.

2.3.2 Sacral Acceleration

Control of motion may differ between the anterior-posterior, mediolateral and vertical directions due to differences in the availability of strategies between directions (Schütte et al., 2015; Winter, 1995). However, the local coordinate system of the sacral IMU may not have aligned with the horizontal-vertical coordinate system of the treadmill. Therefore, the accelerometer readings were transformed to an East-North-Up treadmill-fixed coordinate system. A constant north-pointing magnetometer vector was created to align north with the long axis of the treadmill belt and remove heading drift (Gui et al., 2015). A Kalman filter was applied using the *ahrsfilter* function in MATLAB to estimate IMU orientation by fusing accelerometer, gyroscope and synthetic magnetometer data, enabling transformation of accelerometer readings to a treadmill-fixed coordinate system.

Both sampling frequency and cadence directly influence sample entropy since they determine the spatial distance between adjacent data points and the number of data points within each stride cycle, thus influencing the likelihood of similar vectors being repeated (Ahmadi et al., 2018; Raffalt et al., 2019). Thus, strides were defined by initial contact events that were found using a continuous wavelet transformation with a mother scale = 16 (McCamley et al., 2012). 30 strides (Leverick et al., 2014) of non-wavelet transformed data from the start of each trial were then extracted. These 30 strides were resampled to N = 3000 data points using a cubic spline interpolation to fix the average number of data points per stride. The sample entropy algorithm was then applied to individual components (mediolateral, anterior-posterior, vertical) and the resultant of these resampled time series. Parameters m and r were varied such that all combinations in the respective sets {2, 3, 4} and {0.10, 0.15, 0.20, 0.25, 0.30} were tested (Yentes and Raffalt, 2021).

2.4 Statistical Analysis

Between-day reliability was quantified for each of the six conditions and overall using ICC estimates and their 95% confidence intervals (CIs) based on a single-measure, absolute-agreement, 2-way mixed-effects model. ICC thresholds of < 0.50, 0.50 – 0.75, 0.75 – 0.90 and > 0.90 were used to indicate poor, moderate, good and excellent relative reliability respectively (Koo and Li, 2016). The standard error of measurement (SEM) and absolute minimum detectable change were calculated using the method of Weir (2005) to, respectively, provide an index of absolute reliability and to quantify the minimum change between repeated measurements on an individual that could be considered real.

3. Results

Stride time mean, CV and DFA-α during each trial and day are displayed in Table 2. The running speed and cadence for each participant for each trial are in supplementary material A. Stride time mean displayed excellent relative reliability in all six trials (ICC ≥ 0.935), and overall (ICC = 0.957). Stride time CV displayed varying relative reliability across trials (ICC: 0.441 to 0.783), although with absolute minimum detectable changes ≥ 0.371%. Stride time DFA-α displayed poor relative reliability for trials 2, 3 and 6 and overall (ICC ≤ 0.423) but moderate for trials 1, 4 and 5 (ICC: 529 to 0.562), although with a wide range. Minimum detectable changes were greater than 0.208 across all trials for DFA-α.

Mean sacral acceleration sample entropy values for each combination of the parameters *m* and *r* are shown in Figure 1. Vertical and resultant sacral accelerations displayed the greatest regularity for any fixed pairing of *m* and *r*. ICCs, SEMs and minimum detectable changes for sacral acceleration sample entropy are shown in Figure 2. Mediolateral and resultant accelerations displayed moderate to good relative reliability for all combinations of *m* and *r* (ICC: 0.638 to 0.779 and ICC: 0.641 to 0.798, respectively), whereas anterior-posterior and vertical accelerations displayed poor to moderate relative reliability for all combinations of *m* and *r* (ICC: 0.491 to 0.593 and ICC: 0.420 to 0.610, respectively). The highest overall relative

reliability was found for the combination *m* = 3 and *r* = 0.20 applied to the resultant acceleration signal (ICC = 0.798). This combination of parameters and signal also demonstrated moderate to good relative reliability across all trials (ICC: 0.648 to 0.898), although with relatively high minimum detectable changes (Table 3).

Scatter plots showing the distribution of stride time mean, CV and DFA-α, and resultant sacral acceleration sample entropy for the parameters *m* = 3 and *r* = 0.20, across days 1 and 2, are shown in Figure 3.

4. Discussion

The purpose of this study was to assess the between-day reliability of variability measures in two high-order running gait variables measured using wearable devices. Stride time CV displayed moderate relative reliability (ICC: 0.672), although with relatively high minimum detectable changes. In contrast, stride time DFA-α displayed poor relative and absolute reliability (ICC: 0.457; minimum detectable change: 0.250). Sample entropy when applied to mediolateral and resultant sacral accelerations displayed moderate to good relative reliability for all combinations of the parameters *m* and *r*. However, like stride time CV, it demonstrated poor absolute reliability in the context of between-condition differences observed in this study and changes previously reported within individuals when they transition from healthy to injured states (Gruber et al., 2021).

Stride time mean (day 1: 0.729 s; day 2: 0.732 s), CV (day 1: 1.25%; day 2: 1.21%) and DFA-α (day 1: 0.747; day 2: 0.753) calculated within this study fell within ranges previously reported for treadmill running (Mann et al., 2015; Mo and Chow, 2019) and in overground running studies using IMUs (Brahms et al., 2022; Meardon et al., 2011). Stride time mean displayed excellent between-session reliability across all trials (ICC point estimates: 0.935 to 0.968), whereas stride time CV displayed varying relative reliability across trials (ICC point estimates: 0.441 to 0.783), and with wide ranging CIs. Stride time CV ICC point estimates were similar to those reported by Godin et

al. (2024) (0.52 to 0.62), who also reported wide CIs, but they were lower than those reported by Mo and Chow (2019) and Fuller et al. (2018) who used smaller sample sizes (n = 7 and n = 10 respectively).

In contrast, stride time DFA-α displayed poor to moderate relative reliability across all trials and overall (ICC point estimates: 0.357 to 0.562; minimum detectable changes ≥ 0.208), albeit again with a wide range. Although these point estimates fell within the range reported by Mo and Chow (2019) (ICCs: 0.20 to 0.81), both Godin et al. (2024) (ICCs: 0.79 to 0.85; minimum detectable changes ≤ 0.18) and Fuller et al. (2018) (ICCs: 0.74 to 0.87) found excellent reliability in stride time DFA-α. Moreover, the present study found an minimum detectable change of 0.208 even in the trial that displayed the greatest between-day reliability. Whilst some day-by-day variation in DFA-α would be expected given that the human body is continually self-organising under the presence of dynamic individual constraints (Davids et al., 2003), differences of the magnitude 0.11 to 0.17 have previously been found to separate previously injured and non-injured groups (Meardon et al., 2011), fatigued and non-fatigued states within individuals (Brahms et al., 2022) and the periods before and after fatigue heaving training within individuals (Bellenger et al., 2019). With the DFA algorithm presented here applied to stride time data collected using a RunScribe™ IMU, it would be impossible to confidently interpret within participant changes in DFA-α of an equivalent magnitude as real. However, it is possible that between-day reliability may be different for alternative stride time measurement methodologies, such as optoelectronic devices (Godin et al., 2024), force sensitive resistors (Fuller et al., 2018) or motion capture (Mo and Chow, 2019), or for alternative DFA parameter choices (Godin et al., 2024; Fuller et al., 2018; Mo and Chow, 2019), given that it is highly sensitive to parameter selection (Phinyomark et al., 2020).

Sacral acceleration sample entropy values were in the range 0.168 to 0.819 across trials, with vertical and resultant accelerations displaying the greatest regularity. This may partially be explained by the constant force of gravity in the vertical direction (Schütte et al., 2015). For the

parameters m = 2 and r = 0.20, our sample entropy values (Figure 1) were typically greater than those reported in healthy, unfatigued runners by Schütte et al. (2015). However, Schütte et al. (2015) sampled at 400 Hz, and did not resample their data to fix the average number of data points per stride. Given the resampling procedure used here generated a series with temporal spacing equivalent to sampling rate of 137 Hz (100 points per stride / 0.730 mean stride time), and the decrease in sample entropy with sampling frequency (Raffalt et al., 2019), this difference is reasonable.

To the best of our knowledge, these are the first results on the reliability of sacral acceleration sample entropy during running. ICC point estimates indicated poor to moderate (0.420 to 0.610) relative reliability of anterior-posterior and vertical acceleration signals for all combinations of the parameter m and r, and moderate to good (0.638 to 0.798) relative reliability of mediolateral and resultant sacral acceleration signals. Greater deficits in mediolateral control compared to other directions have been attributed to the fewer available degrees of freedom to control motion in this direction (Morrison et al., 2016), and the ICC results observed in this study would support the long-term monitoring of mediolateral sacral acceleration signals in running, particularly if changes in mediolateral sample entropy due to reduced degrees of freedom emerge earlier than in other directions. However, SEM and minimum detectable change values indicated the limitations of interpreting between-day changes in sample entropy as a change in human health. For the parameters m = 2 and r = 0.15 used by Gruber et al. (2021) in their proof-of-concept study on the relationship between sample entropy and running-related injury, we found an minimum detectable change of 0.143 for resultant sacral acceleration. This was over 46 times greater than the group average difference they reported in sample entropy between measurements at baseline and immediately prior to running-related injury. Moreover, our minimum detectable change values also scaled approximately linearly with sample entropy values as the parameters m and r varied. Thus, the reliability of sample entropy as presented here may be inadequate to interpret

changes in sample entropy as real independent of parameter selection, unless greater within participant changes are observed.

This study had limitations that must be considered. First, DFA was performed with a single set of input parameters, although this was selected following a systematic process for a series of constant length (N = 512 strides). Since the implementation of DFA in running gait literature varies widely (Wilson and Likens, 2023), future studies should investigate the change in reliability for alternative parameter selections. Second, our study was limited to fixed speed indoor treadmill running so that the effects of changes in running speed and in environmental conditions, which may have reduced between-day reliability, were minimised. Further studies are needed to understand repeatability of variability measures in representative overground running trials. Third, stride time data were calculated using proprietary algorithms from RunScribe™ which may apply a level of smoothing that removes some variability. Alternative stride time measurement methodologies may yield different results. Finally, between-day differences likely reflected a combination of measurement error and natural day-to-day biological variation that could not be separated. This could be further investigated by use of gold standard laboratory measurement systems.

5. Conclusion

This study investigated the between-day reliability of variability measures in two high-order variables measured using IMUs. Sample entropy of sacrum acceleration signals displayed good relative reliability for mediolateral and resultant acceleration across a range of speed and gradient combinations, although with minimum detectable changes larger than between-day changes reported in previous studies. In contrast, stride time data from RunScribe™ IMUs programmed with proprietary algorithms yielded CV values displaying highly variable reliability

and poor reliability for DFA-α. Any between-day changes in variability measures of stride time or sacral acceleration collected using an IMU should thus be interpreted with caution.

Acknowledgements:

None.

Funding:

This work received no external funding. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version of this paper arising from this submission.

Declaration of Competing Interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A: Supplementary material

References

Ahmadi, S., Sepehri, N., Wu, C., Szturm, T., 2018. Sample Entropy of Human Gait Center of Pressure Displacement: A Systematic Methodological Analysis. Entropy 20, 579. https://doi.org/10.3390/e20080579

Bellenger, C.R., Arnold, J.B., Buckley, J.D., Thewlis, D., Fuller, J.T., 2019. Detrended fluctuation analysis detects altered coordination of running gait in athletes following a heavy period of training. Journal of Science and Medicine in Sport 22, 294–299. https://doi.org/10.1016/j.jsams.2018.09.002

Brahms, C.M., Zhao, Y., Gerhard, D., Barden, J.M., 2022. Long-range correlations and stride pattern variability in recreational and elite distance runners during a prolonged run. Gait & Posture 92, 487–492. https://doi.org/10.1016/j.gaitpost.2020.08.107

Dai, S., Carroll, D.D., Watson, K.B., Paul, P., Carlson, S.A., Fulton, J.E., 2015. Participation in Types of Physical Activities among US Adults—National Health and Nutrition Examination Survey 1999–2006. Journal of Physical Activity and Health 12, S128–S140. https://doi.org/10.1123/jpah.2015-0038

Davids, K., Glazier, P., Araújo, D., Bartlett, R., 2003. Movement Systems as Dynamical Systems. Sports Med 33, 245–260. https://doi.org/10.2165/00007256-200333040-00001

Delignières, D., 2015. Correlation Properties of (Discrete) Fractional Gaussian Noise and Fractional Brownian Motion. Mathematical Problems in Engineering 2015, e485623. https://doi.org/10.1155/2015/485623

Di Bacco, V.E., Gage, W.H., 2023. Evaluation of a smartphone accelerometer system for measuring nonlinear dynamics during treadmill walking: Concurrent validity and test-retest reliability. Journal of Biomechanics 151, 111527. https://doi.org/10.1016/j.jbiomech.2023.111527

Fonseca, S.T., Souza, T.R., Verhagen, E., Van Emmerik, R., Bittencourt, N.F.N., Mendonça, L.D.M., Andrade, A.G.P., Resende, R.A., Ocarino, J.M., 2020. Sports Injury Forecasting and Complexity: A Synergetic Approach. Sports Med 50, 1757–1770. https://doi.org/10.1007/s40279-020-01326-4

Fuller, J., Buckley, J., Tsiros, M., Thewlis, D., 2018. Reliability of the long-range correlations obtained from detrended fluctuation analysis of running stride intervals: 36th Conference of the International Society of Biomechanics in Sports, ISBS 2018. International Society of Biomechanics in Sports Proceedings Archive 36, 702–705.

Godin, A., Rouget, L., Eustache, E., Mourot, L., Sagawa, Y., 2024. Evaluation of the optimal number of steps to obtain reliable running spatio-temporal parameters and their variability. Gait & Posture 111, 37–43. https://doi.org/10.1016/j.gaitpost.2024.04.012

Gruber, A.H., McDonnell, J., Davis, J.J., Vollmar, J.E., Harezlak, J., Paquette, M.R., 2021. Monitoring Gait Complexity as an Indicator for Running-Related Injury Risk in Collegiate Cross-Country Runners: A Proof-of-Concept Study. Front. Sports Act. Living 3, 630975. https://doi.org/10.3389/fspor.2021.630975

Gui, P., Tang, L., Mukhopadhyay, S., 2015. MEMS based IMU for tilting measurement: Comparison of complementary and kalman filter based data fusion, in: 2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA). Presented at the 2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA), IEEE, Auckland, New Zealand, pp. 2004–2009. https://doi.org/10.1109/ICIEA.2015.7334442

Hansen, E.A., Nielsen, A.M., Kristensen, L.A.R., Madeleine, P., Voigt, M., 2018. Prediction of walk-to-run transition using stride frequency: A test-retest reliability study. Gait & Posture 60, 71–75. https://doi.org/10.1016/j.gaitpost.2017.11.008

Jordan, K., Challis, J.H., Newell, K.M., 2007. Speed influences on the scaling behavior of gait cycle fluctuations during treadmill running. Human Movement Science 26, 87–102. https://doi.org/10.1016/j.humov.2006.10.001

Jordan, K., Challis, J.H., Newell, K.M., 2006. Long range correlations in the stride interval of running. Gait & Posture 24, 120–125. https://doi.org/10.1016/j.gaitpost.2005.08.003
Kemler, E., Blokland, D., Backx, F., Huisstede, B., 2018. Differences in injury risk and characteristics of injuries between novice and experienced runners over a 4-year period. The Physician and Sportsmedicine 46, 485–491. https://doi.org/10.1080/00913847.2018.1507410
Koldenhoven, R.M., Hertel, J., 2018. Validation of a Wearable Sensor for Measuring Running Biomechanics. Digital Biomarkers 2, 74–78. https://doi.org/10.1159/000491645
Koo, T.K., Li, M.Y., 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 15, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012
Kozinc, Ž., Smajla, D., Šarabon, N., 2022. The reliability of wearable commercial sensors for outdoor assessment of running biomechanics: the effect of surface and running speed. Sports Biomechanics 0, 1–14. https://doi.org/10.1080/14763141.2021.2022746
Kroese, D.P., Botev, Z.I., 2013. Spatial Process Generation. https://doi.org/10.48550/arXiv.1308.0399
Lee, D., Brellenthin, A.G., Thompson, P.D., Sui, X., Lee, I.-M., Lavie, C.J., 2017. Running as a Key Lifestyle Medicine for Longevity. Progress in Cardiovascular Diseases 60, 45–55. https://doi.org/10.1016/j.pcad.2017.03.005
Leverick, G., Szturm, T., Wu, C.Q., 2014. Using Entropy Measures to Characterize Human Locomotion. Journal of Biomechanical Engineering 136. https://doi.org/10.1115/1.4028410
Liddy, J.J., Haddad, J.M., 2018. Evenly spaced Detrended Fluctuation Analysis: Selecting the number of points for the diffusion plot. Physica A: Statistical Mechanics and its Applications 491, 233–248. https://doi.org/10.1016/j.physa.2017.08.099
Mann, R., Malisoux, L., Nührenbörger, C., Urhausen, A., Meijer, K., Theisen, D., 2015. Association of previous injury and speed with running style and stride-to-stride fluctuations. Scand J Med Sci Sports 25. https://doi.org/10.1111/sms.12397
McCamley, J., Donati, M., Grimpampi, E., Mazzà, C., 2012. An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data. Gait & Posture 36, 316–318. https://doi.org/10.1016/j.gaitpost.2012.02.019
Meardon, S.A., Hamill, J., Derrick, T.R., 2011. Running injury and stride time variability over a prolonged run. Gait & Posture 33, 36–40. https://doi.org/10.1016/j.gaitpost.2010.09.020
Mo, S., Chow, D.H.K., 2019. Reliability of the fluctuations within the stride time series measured in runners during treadmill running to exhaustion. Gait & Posture 74, 1–6. https://doi.org/10.1016/j.gaitpost.2019.08.008
Mo, S., Chow, D.H.K., 2018. Stride-to-stride variability and complexity between novice and experienced runners during a prolonged run at anaerobic threshold speed. Gait & Posture 64, 7–11. https://doi.org/10.1016/j.gaitpost.2018.05.021
Morrison, S., Rynders, C.A., Sosnoff, J.J., 2016. Deficits in medio-lateral balance control and the implications for falls in individuals with multiple sclerosis. Gait & Posture 49, 148–154. https://doi.org/10.1016/j.gaitpost.2016.06.036
Newell, K.M., Vaillancourt, D.E., 2001. Dimensional change in motor learning. Human Movement Science 20, 695–715. https://doi.org/10.1016/S0167-9457(01)00073-2
Pavlova, O.N., Abdurashitov, A.S., Ulanova, M.V., Shushunova, N.A., Pavlov, A.N., 2019. Effects of missing data on characterization of complex dynamics from time series. Communications in Nonlinear Science and Numerical Simulation 66, 31–40. https://doi.org/10.1016/j.cnsns.2018.06.002
Peng, C.K., Havlin, S., Stanley, H.E., Goldberger, A.L., 1995. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 5, 82–87. https://doi.org/10.1063/1.166141

Phinyomark, A., Larracy, R., Scheme, E., 2020. Fractal Analysis of Human Gait Variability via Stride Interval Time Series. Front. Physiol. 11, 333. https://doi.org/10.3389/fphys.2020.00333

Quirino, J., Santos, T.R.T., Okai-Nóbrega, L.A., De Araújo, P.A., Carvalho, R., Ocarino, J.D.M., Souza, T.R., Fonseca, S.T., 2021. Runners with a history of injury have greater lower limb movement regularity than runners without a history of injury. Sports Biomechanics 1–13. https://doi.org/10.1080/14763141.2021.1929435

Raffalt, P.C., Alkjær, T., Brynjólfsson, B., Jørgensen, L., Bartholdy, C., Henriksen, M., 2018. Day-to-Day Reliability of Nonlinear Methods to Assess Walking Dynamics. Journal of Biomechanical Engineering 140. https://doi.org/10.1115/1.4041044

Raffalt, P.C., McCamley, J., Denton, W., Yentes, J.M., 2019. Sampling frequency influences sample entropy of kinematics during walking. Med Biol Eng Comput 57, 759–764. https://doi.org/10.1007/s11517-018-1920-2

Ravi, D.K., Marmelat, V., Taylor, W.R., Newell, K.M., Stergiou, N., Singh, N.B., 2020. Assessing the Temporal Organization of Walking Variability: A Systematic Review and Consensus Guidelines on Detrended Fluctuation Analysis. Frontiers in Physiology 11.

Richman, J.S., Moorman, J.R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology 278, H2039–H2049. https://doi.org/10.1152/ajpheart.2000.278.6.H2039

Ryan, N.S., Bruno, P.A., Barden, J.M., 2020. Test–Retest Reliability and the Effects of Walking Speed on Stride Time Variability During Continuous, Overground Walking in Healthy Young Adults. Journal of Applied Biomechanics 37, 102–108. https://doi.org/10.1123/jab.2020-0138

Schütte, K.H., Maas, E.A., Exadaktylos, V., Berckmans, D., Venter, R.E., Vanwanseele, B., 2015. Wireless Tri-Axial Trunk Accelerometry Detects Deviations in Dynamic Center of Mass Motion Due to Running-Induced Fatigue. PLOS ONE 10, e0141957. https://doi.org/10.1371/journal.pone.0141957

Schütte, K.H., Seerden, S., Venter, R., Vanwanseele, B., 2018. Influence of outdoor running fatigue and medial tibial stress syndrome on accelerometer-based loading and stability. Gait & Posture 59, 222–228. https://doi.org/10.1016/j.gaitpost.2017.10.021

Stergiou, N., Decker, L.M., 2011. Human movement variability, nonlinear dynamics, and pathology: Is there a connection? Human Movement Science, EWOMS 2009: The European Workshop on Movement Science 30, 869–888. https://doi.org/10.1016/j.humov.2011.06.002

Tudor-Locke, C., Johnson, W.D., Katzmarzyk, P.T., 2010. Frequently Reported Activities by Intensity for U.S. Adults. American Journal of Preventive Medicine 39, e13–e20. https://doi.org/10.1016/j.amepre.2010.05.017

Ueno, H., Nakazawa, S., Takeuchi, Y., Sugita, M., 2021. Relationship between Step Characteristics and Race Performance during 5000-m Race. Sports (Basel) 9, 131. https://doi.org/10.3390/sports9090131

Van Hooren, B., Fuller, J.T., Buckley, J.D., Miller, J.R., Sewell, K., Rao, G., Barton, C., Bishop, C., Willy, R.W., 2020. Is Motorized Treadmill Running Biomechanically Comparable to Overground Running? A Systematic Review and Meta-Analysis of Cross-Over Studies. Sports Med 50, 785–813. https://doi.org/10.1007/s40279-019-01237-z

Videbæk, S., Bueno, A.M., Nielsen, R.O., Rasmussen, S., 2015. Incidence of Running-Related Injuries Per 1000 h of running in Different Types of Runners: A Systematic Review and Meta-Analysis. Sports Med 45, 1017–1026. https://doi.org/10.1007/s40279-015-0333-8

Weir, J., 2005. Quantifying Test-Retest Reliability Using The Intraclass Correlation Coefficient and the SEM. Journal of strength and conditioning research / National Strength & Conditioning Association 19, 231–40. https://doi.org/10.1519/15184.1

Willy, R.W., 2018. Innovations and pitfalls in the use of wearable devices in the prevention and rehabilitation of running related injuries. Physical Therapy in Sport 29, 26–33. https://doi.org/10.1016/j.ptsp.2017.10.003

Wilson, T.J., Likens, A.D., 2023. Running gait produces long range correlations: A systematic review. Gait & Posture 102, 171–179. https://doi.org/10.1016/j.gaitpost.2023.04.001

Wilson, T.J., Mangalam, M., Stergiou, N., Likens, A.D., 2023. Multifractality in stride-to-stride variations reveals that walking involves more movement tuning and adjusting than running. Front. Netw. Physiol. 3. https://doi.org/10.3389/fnetp.2023.1294545

Winter, D., 1995. Human balance and posture control during standing and walking. Gait & Posture 3, 193–214. https://doi.org/10.1016/0966-6362(96)82849-9

Worp, M.P. van der, Haaf, D.S.M. ten, Cingel, R. van, Wijer, A. de, Sanden, M.W.G.N. der, Staal, J.B., 2015. Injuries in Runners; A Systematic Review on Risk Factors and Sex Differences. PLOS ONE 10, e0114937. https://doi.org/10.1371/journal.pone.0114937

Yentes, J.M., Raffalt, P.C., 2021. Entropy analysis in gait research – methodological considerations and recommendations. Ann Biomed Eng 49, 979–990. https://doi.org/10.1007/s10439-020-02616-8

Table 1: The six ordered conditions participants completed during each day. 5 km speeds were self-reported based on participants' current level of fitness.

Trial Order	Trial Description
1	Preferred running speed at 0% gradient
2	75% 5 km speed at 0% gradient
3	85% 5 km speed at 0% gradient
4	95% 5 km speed at 0% gradient
5	75% 5 km speed at -5% gradient
6	75% 5 km speed at +5% gradient

Table 2: Between-day reliability of stride time mean, CV and DFA-α split by trial.

	Trial	Condition	Day 1 Mean (95% CI)	Day 2 Mean (95% CI)	ICC (95% CI)	SEM	MDC
Stride time Mean (s)	1	PRS at 0% gradient	0.736 (0.713, 0.759)	0.736 (0.713, 0.759)	0.935 (0.840, 0.975)	0.011	0.032
	2	75% 5 km speed at 0% gradient	0.739 (0.717, 0.762)	0.743 (0.717, 0.762)	0.939 (0.853, 0.976)	0.011	0.031
	3	85% 5 km speed at 0% gradient	0.724 (0.701, 0.748)	0.726 (0.701, 0.748)	0.961 (0.904, 0.985)	0.009	0.025
	4	95% 5 km speed at 0% gradient	0.704 (0.680, 0.728)	0.709 (0.680, 0.728)	0.959 (0.896, 0.984)	0.010	0.027
	5	75% 5 km speed at -5% gradient	0.750 (0.728, 0.773)	0.751 (0.728, 0.773)	0.966 (0.914, 0.987)	0.008	0.023
	6	75% 5 km speed at +5% gradient	0.723 (0.701, 0.744)	0.726 (0.701, 0.744)	0.968 (0.919, 0.987)	0.008	0.022
		Overall	0.729 (0.720, 0.738)	0.732 (0.723, 0.741)	0.957 (0.938, 0.970)	0.010	0.027
Stride time CV (%)	1	PRS at 0% gradient	1.325 (1.109, 1.540)	1.217 (1.109, 1.540)	0.675 (0.342, 0.860)	0.211	0.585
	2	75% 5 km speed at 0% gradient	1.249 (1.048, 1.449)	1.133 (1.048, 1.449)	0.441 (0.022, 0.735)	0.249	0.690
	3	85% 5 km speed at 0% gradient	1.202 (1.020, 1.385)	1.198 (1.020, 1.385)	0.753 (0.460, 0.897)	0.166	0.460
	4	95% 5 km speed at 0% gradient	1.272 (1.093, 1.451)	1.284 (1.093, 1.451)	0.716 (0.394, 0.881)	0.193	0.535
	5	75% 5 km speed at -5% gradient	1.246 (1.114, 1.378)	1.285 (1.114, 1.378)	0.743 (0.451, 0.892)	0.149	0.414
	6	75% 5 km speed at +5% gradient	1.205 (1.048, 1.361)	1.169 (1.048, 1.361)	0.783 (0.525, 0.910)	0.134	0.371
		Overall	1.250 (1.182, 1.318)	1.214 (1.161, 1.268)	0.672 (0.557, 0.761)	0.189	0.525
Stride time DFA-α	1	PRS at 0% gradient	0.703 (0.642, 0.764)	0.775 (0.642, 0.764)	0.562 (0.098, 0.814)	0.081	0.226
	2	75% 5 km speed at 0% gradient	0.751 (0.688, 0.815)	0.728 (0.688, 0.815)	0.357 (-0.109, 0.692)	0.103	0.287
	3	85% 5 km speed at 0% gradient	0.760 (0.702, 0.818)	0.754 (0.702, 0.818)	0.423 (-0.043, 0.733)	0.092	0.255
	4	95% 5 km speed at 0% gradient	0.746 (0.678, 0.814)	0.744 (0.678, 0.814)	0.562 (0.145, 0.806)	0.100	0.276
	5	75% 5 km speed at -5% gradient	0.741 (0.697, 0.785)	0.763 (0.697, 0.785)	0.529 (0.116, 0.787)	0.075	0.208
	6	75% 5 km speed at +5% gradient	0.779 (0.729, 0.828)	0.752 (0.729, 0.828)	0.347 (-0.107, 0.684)	0.081	0.225
		Overall	0.747 (0.724, 0.769)	0.753 (0.729, 0.776)	0.457 (0.299, 0.591)	0.090	0.250

CV: coefficient of variation; DFA: detrended fluctuation analysis; PRS: preferred running speed; CI: confidence interval; ICC: intraclass correlation coefficient; SEM: standard error of measurement; MDC: minimum detectable change.

Table 3: Between-day reliability of resultant sacral acceleration sample entropy. Results are presented for the combination of the parameters *m* and *r* that yielded the highest ICC overall (*m* = 3, *r* = 0.20).

	Trial	Condition	Day 1 Mean (95% CI)	Day 2 Mean (95% CI)	ICC (95% CI)	SEM	MDC
Sample entropy	1	PRS at 0% gradient	0.265 (0.229, 0.302)	0.256 (0.229, 0.302)	0.856 (0.670, 0.942)	0.028	0.079
	2	75% 5 km speed at 0% gradient	0.270 (0.234, 0.306)	0.262 (0.234, 0.306)	0.745 (0.451, 0.893)	0.036	0.101
	3	85% 5 km speed at 0% gradient	0.283 (0.247, 0.319)	0.283 (0.247, 0.319)	0.898 (0.755, 0.960)	0.025	0.069
	4	95% 5 km speed at 0% gradient	0.298 (0.257, 0.340)	0.299 (0.257, 0.340)	0.830 (0.610, 0.931)	0.037	0.101
	5	75% 5 km speed at -5% gradient	0.264 (0.229, 0.299)	0.258 (0.229, 0.299)	0.757 (0.471, 0.899)	0.037	0.103
	6	75% 5 km speed at +5% gradient	0.270 (0.243, 0.297)	0.281 (0.243, 0.297)	0.648 (0.293, 0.847)	0.039	0.109
		Overall	0.275 (0.261, 0.289)	0.273 (0.258, 0.288)	0.798 (0.720, 0.856)	0.034	0.095

PRS: preferred running speed; CI: confidence interval; ICC: intraclass correlation coefficient; SEM: standard error of measurement; MDC: minimum detectable change.

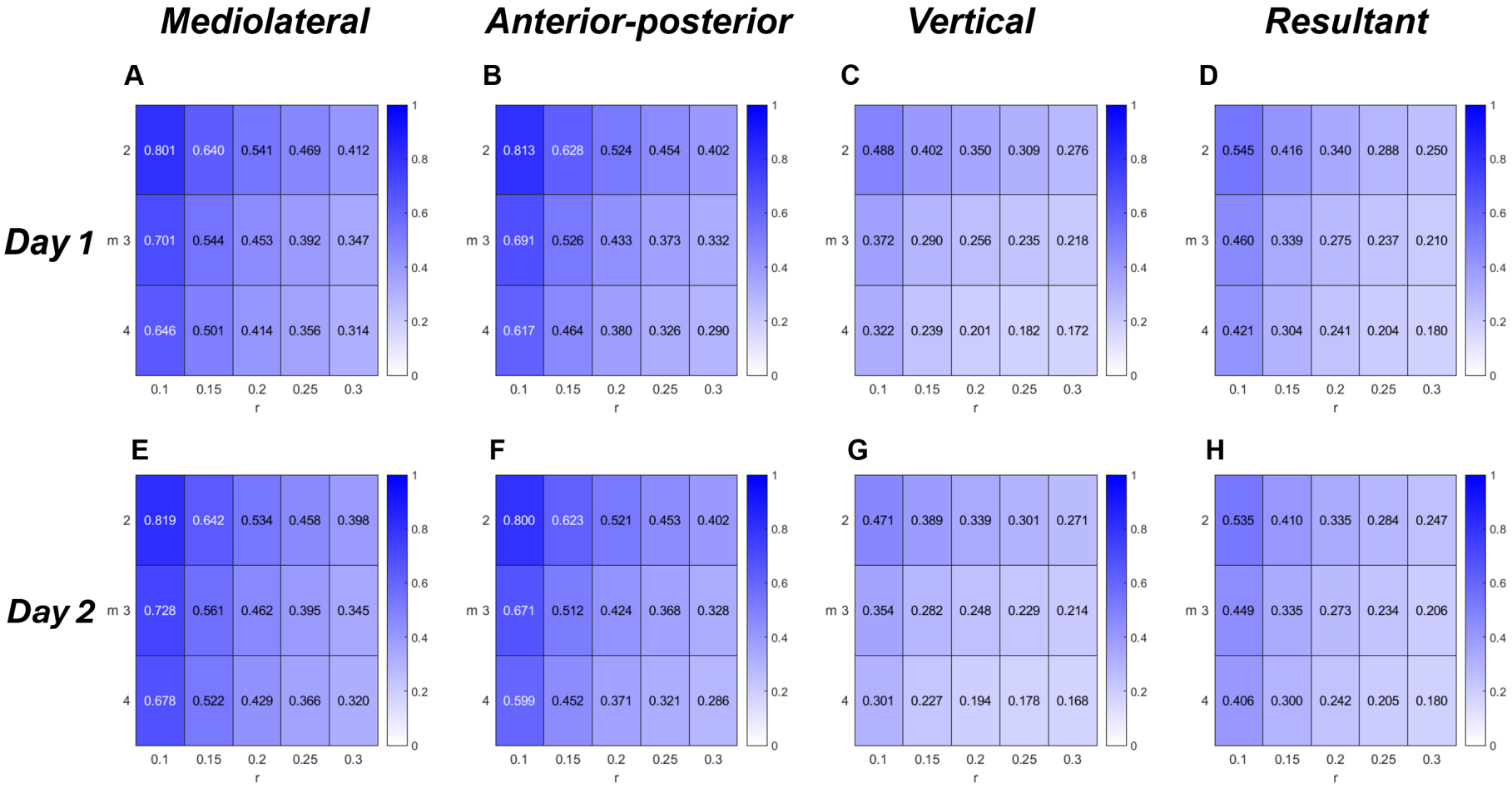

Figure 1: Mean sample entropy values for each combination of the tested parameters m and r during day 1 for sacral A) mediolateral acceleration, B) anterior-posterior acceleration, C) vertical acceleration, D) resultant acceleration, and during day 2 for sacral E) mediolateral acceleration, F) anterior-posterior acceleration, G) vertical acceleration, H) resultant acceleration.

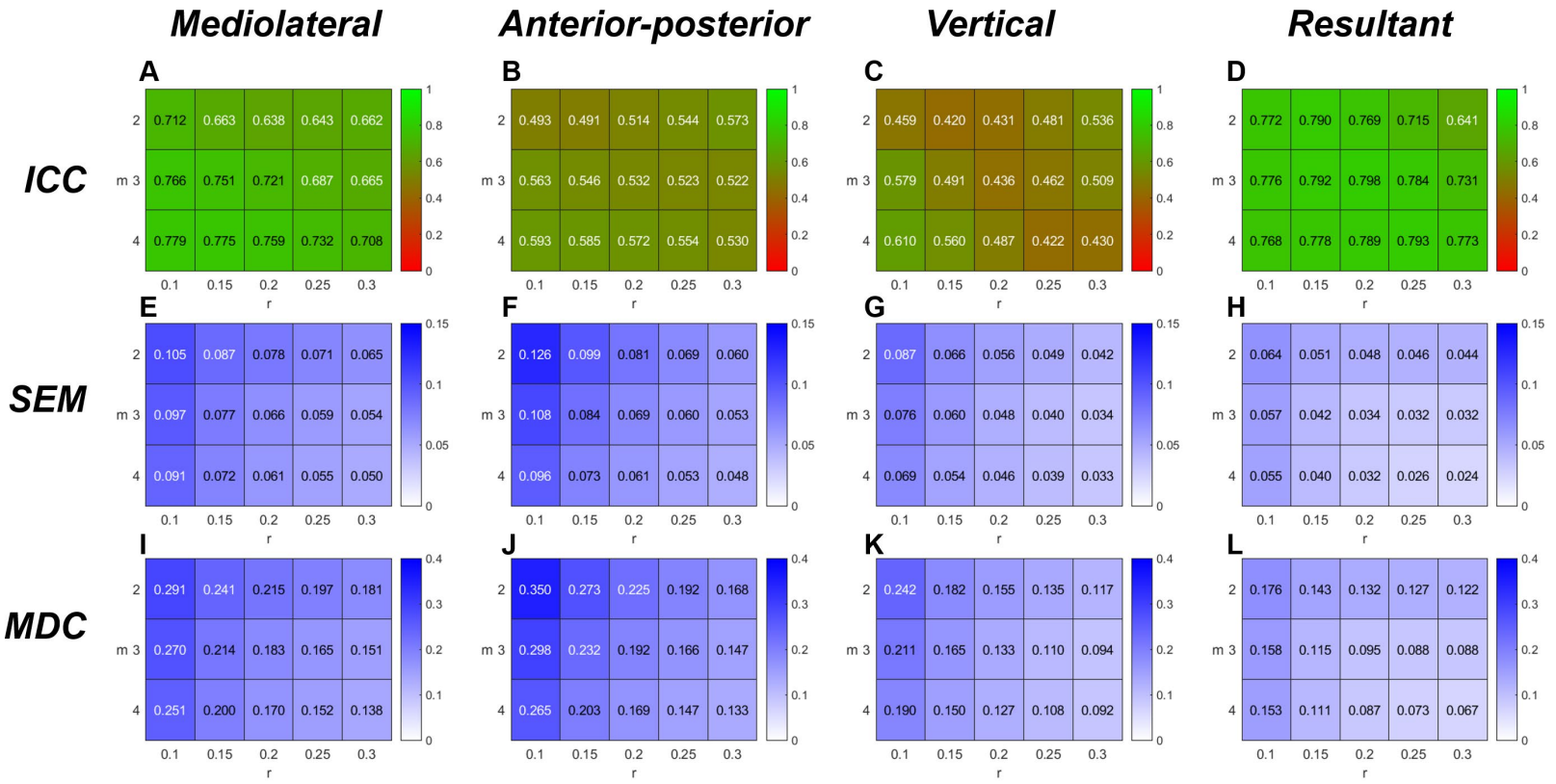

Figure 2: Between-day reliability of sacral acceleration signals. Top row: ICC estimates for each combination of the tested parameters m and r for sacral A) mediolateral acceleration, B) anterior-posterior acceleration, C) vertical acceleration, D) resultant acceleration. Middle row: SEM for each combination of the tested parameters m and r for sacral E) mediolateral acceleration, F) anterior-posterior acceleration, G) vertical acceleration, H) resultant acceleration. Bottom row: MDC for each combination of the tested parameters m and r for sacral I) mediolateral acceleration, J) anterior-posterior acceleration, K) vertical acceleration, L) resultant acceleration.

ICC: intraclass correlation coefficient; SEM: standard error of measurement; MDC: minimum detectable change.

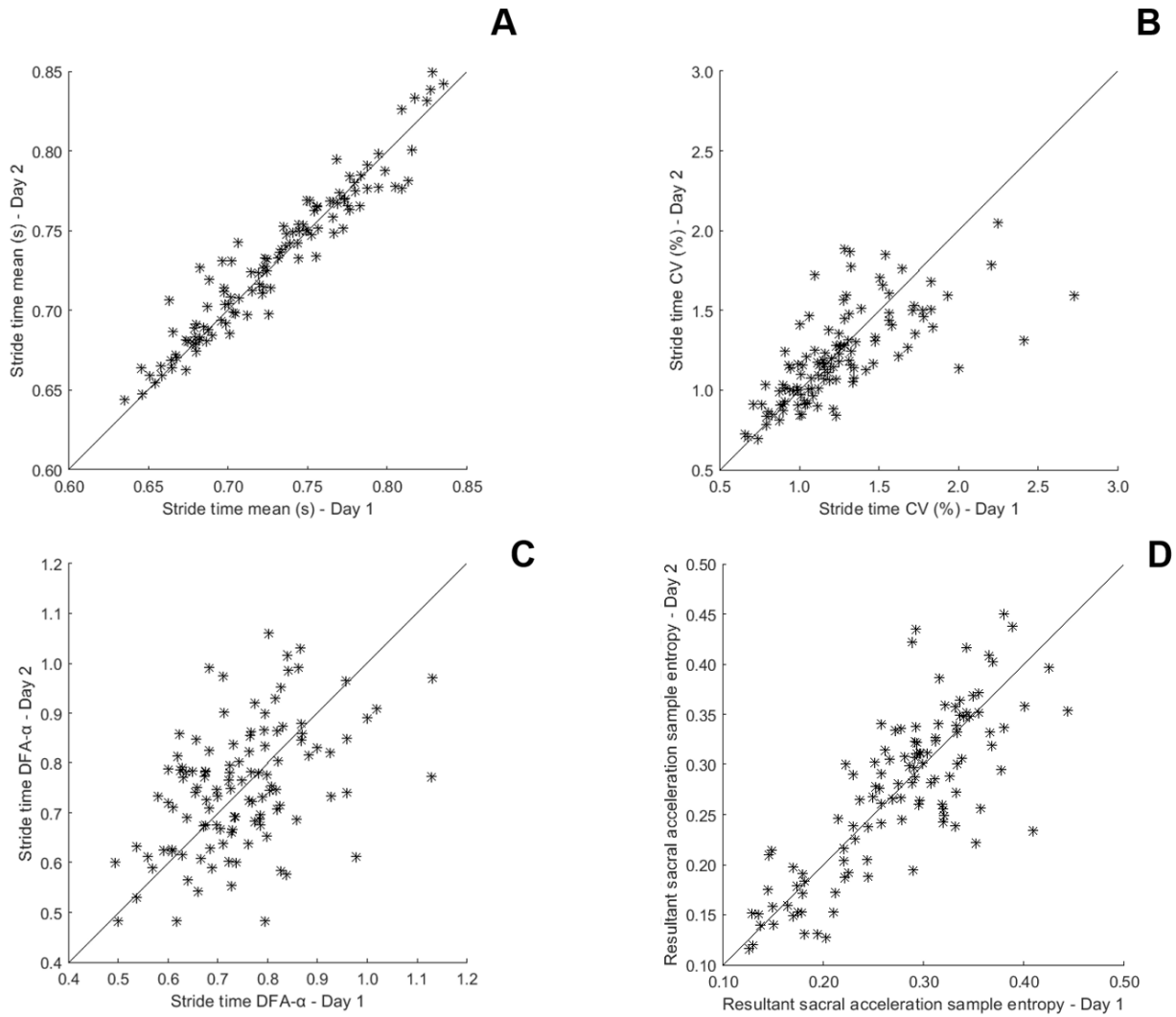

Figure 3: Distribution of A) stride time mean, B) stride time coefficient of variation (CV), C) stride time detrended fluctuation analysis exponent alpha (DFA-α), and D) resultant sacral acceleration sample entropy for the parameters m = 3 and r = 0.20, across days 1 and 2. Values fall on the solid black line if they are identical on days 1 and 2.