Computers in Biology and Medicine 119 (2020) 103673

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

Influence of sampling frequency and number of strides on recurrence quantifiers extracted from gait data

Alfredo de Oliveira Assis [a,**], Fábio Barbosa Rodrigues [a], Adriano Carafini [a], Thiago Santana Lemes [a], Guilherme Augusto Gomes de Villa [a], Adriano de Oliveira Andrade [b], Marcus Fraga Vieira [a]

[a] Bioengineering and Biomechanics Laboratory, Federal University of Goiás, Goiânia, Brazil
[b] Centre for Innovation and Technology Assessment in Health, Federal University of Uberlândia, Uberlândia, Brazil

ARTICLE INFO

Keywords:
Recurrence quantification analysis
Gait
Nonlinear analysis
Strides
Sampling frequency

ABSTRACT

In this study, the influence of the sampling frequency and number of strides on recurrence quantifiers extracted from gait data was investigated in order to provide baseline values and preserve the system's non-linear dynamical characteristics expressed by these recurrence quantifiers. Recurrence quantifiers were extracted from a recurrence plot (RP), which required the reconstruction of a high-dimensional state space capable of reproducing the dynamical characteristics of the analyzed system. In this study, the following quantifiers were extracted: rate of recurrence (RR), determinism (DET), average diagonal lines length (AVG), maximum diagonal lines length (MaxL), Shannon entropy (EntD), and measure of trend (TREND). Data collected during treadmill walking were statistically analyzed to compare the distribution characteristics (mean, median, and standard deviation) and the quantifiers' correlation with those obtained from a control time series with an acquisition time corresponding to 150 strides and a 100-Hz sampling frequency, which are common values used in gait studies. It was not possible to reduce the number of strides for the MaxL or TREND. However, for the RR, DET, AVG, and EntD, it was possible to reduce the number of strides by 60% when analyzed together. The minimum sampling frequency required to extract all quantifiers simultaneously was 100 Hz. This potential reduction in the number of strides is appropriate for evaluating fast gait events, with short temporal localization in the RP, by applying the sliding window method to the recurrence plot.

1. Introduction

Gait is defined as the interaction among numerous actions performed by the musculoskeletal system and postural control with the aim of maintaining the overall system's balance and stability [1]. These interactions result in specific and dynamical characteristics for each person, such as stability and determinism [2,3].

A number of nonlinear features of gait stability and variability have been extensively used for identifying pathologies and frailties in different populations as well as assessing fall risk and evaluating rehabilitation procedures [2,4–15]. These features have shown to be sensitive to several parameters related to data acquisition, such as sample size, sampling frequency, and type of experiment [16]. In some cases, the calculation of these features requires a longer time series to provide robust results [17–20].

When dealing with some populations, such as older adults, impaired individuals, neurological patients, and people suffering from certain pathologies, several issues arise from prolonged and exhausting data acquisition. Besides the inherent discomfort experienced by the participants, the results of the analysis may also be affected due to the difficulty that some individuals have in maintaining the protocol directives for a long period [21,22]. Thus, it would be beneficial to reduce the data collection time.

Recurrence quantification analysis (RQA)—a tool for nonlinear features extraction from a system—has gained attention from researchers who are interested in dynamical analysis of biological phenomena [23,24]. RQA was introduced to objectively quantify recurrence plot (RP) structures. RPs are a way to visualize recurrences in the dynamics of a dynamical system, and they represent the times at which some states recur in a phase space [25].

* Corresponding author.
E-mail address: alfredo.assis@ifg.edu.br (A. de Oliveira Assis).

https://doi.org/10.1016/j.compbiomed.2020.103673
Received 22 August 2019; Received in revised form 22 February 2020; Accepted 22 February 2020
Available online 26 February 2020

The main advantage of RQA is its robustness to contamination that may affect the outcomes yielded by other nonlinear time series analyses [25,26]. Furthermore, other nonlinear techniques—such as the Lyapunov exponent, orbital stability, and entropy—demand long-lasting time series and/or strict periodicity [17–20].

The use of RQA in gait analysis may provide useful information about the pattern and structure of gait dynamics, i.e., how a gait signal evolves over time along strides (gait cycles). Its increasing use is motivated by the fact that several RQA quantifiers require a smaller data length when compared to other nonlinear analyses [14,27].

[27] addressed the influence of the number of strides on the values of certain RQA quantifiers. These can range from a lower limit of 10 strides to an upper limit of 150 strides. This upper limit of 150 strides has also been used in the literature in the calculation of other nonlinear measures, such as the Lyapunov exponent, orbital stability, and entropy [17–20,27].

Another important factor to be considered in RQA is the sampling frequency, which has presented a large range (between 50 Hz and 128 Hz) in most studies in which RQA was applied to gait data [4,9,15,27]. [28]; for example, showed that the Shannon entropy of the RQA is sensitive to the sampling frequency, presenting a particularity that underscores a research need.

Studies that have verified the possibility of a reduction in sampling (sampling frequency) and duration (number of strides) of the signals for nonlinear quantifiers extraction usually evaluated the intraclass correlation (ICC) to establish the reliability of the quantifiers [17–20,27]. Although ICC is sensitive to variations in the mean, standard deviation, and correlation of the quantifier outcomes, it is not possible to determine how the reductions in sampling frequency and number of strides cause deterioration of the quantifiers in these aspects separately [29]. Reducing the number of strides or the sampling frequency of the time series can result in changes in the means and medians (translation of data) or standard deviations (dispersion of data) or in a deterioration of correlations (modification of within-group relations). In this regard, it is necessary to adopt another approach, as in this study, to verify how the quantifiers are altered when extracted from time series with fewer strides and with reduced sampling frequencies compared to a control (e.g., time series with 150 strides and sampling frequency of 100 Hz). Regarding RQA, in addition to relying exclusively on interquartile range and median values, previous studies only analyzed the effect of reducing the number of strides [27].

Therefore, the objective of this study is to analyze the effects of a reduction in the number of strides and sampling frequency of gait data on the RQA quantifiers in order to reduce data collection time and processing time, respectively. The quantifiers analyzed here, namely rate of recurrence (RR), determinism (DET), average diagonal lines length (AVG), maximum diagonal lines length (MaxL), Shannon entropy (EntD), and measure of trend (TREND), have been previously defined in other studies [2,25,27,30].

We hypothesize that quantifiers related to rates and proportions (which are normalized by the number of recurrences), such as RR, DET, AVG, and EntD, would be less sensitive to the reduction in the number of strides and sampling frequency, presenting values that maintain the same proportion within the sample (high correlation) when comparing results extracted from time series with different numbers of strides and sampling frequencies. For the MaxL quantifier, defined as the length of the longest diagonal within the RP, we expect a greater dependence on the number of recurrences and, hence, the number of strides and sampling frequency. Finally, for the TREND quantifier, which is related to the signal stationarity, we expect little dependence on the number of strides and sampling frequency, although TREND can be dependent on the signal size, presenting contrasting results for different data lengths [31].

2. Methods

2.1. Subjects

Data from 90 healthy young adults (44 males and 46 females; age: 24.5 ± 4.9 years; weight: 66.5 ± 12.1 kg; height: 1.71 ± 0.07 m) from previous studies were used in this research [32–34]. In addition, data from 37 healthy older adults from a previous study were analyzed (15 males and 22 females; 69.3 ± 6.6 years old; 66.5 ± 12.0 kg; 1.58 ± 0.08 m) [33], and their results are presented in Supplementary File 1. In the analyzed data, all subjects were healthy with no musculoskeletal injury or pain at the time of data collection. Prior to participation, they voluntarily signed an informed consent form approved by the Institutional Ethics Committee for Human Research.

2.2. Equipment

Gait kinematics data were collected using a 3D motion capture system (Vicon Nexus, Oxford Metrics, Oxford, UK) consisting of 10 infrared cameras operating at 100 Hz. For all subjects, reflective markers were attached to both lateral malleoli, the heels, and the head of the second and fifth metatarsals for step detection and to the spinous process of the T1 vertebrae for measurement of trunk movement and RQA quantifiers extraction, as maintaining stability of the upper body is critical for human gait [35,36].

2.3. Protocol

The subjects walked on a level treadmill (EVO 4000 model, Evolution Inc., Joinville, SC; 0.1 km/h of resolution) wearing their own regular shoes and a safety harness that allowed for natural arm swing. First, anthropometric data (mass and height) were collected using a digital balance and stadiometer (W 200/5 model, Welmy Inc., Santa Bárbara D'Oeste, SP; precision of 0.050 kg and 1 mm, respectively). Initially, after 2 min of familiarization walking on the treadmill, the individual's preferred self-selected walking speed (PWS) was evaluated following a previously reported protocol [37]. Briefly, the speed of the treadmill was gradually increased and then decreased above and below the comfortable walking speed as reported by the subjects. The participants of the study were not aware of the speed of the treadmill and the average of the speed reported as comfortable was computed as the PWS. The average PWS was 1.05 ± 0.12 m/s and 0.84 ± 0.10 m/s for younger and older adults, respectively.

Next, the reflective markers were attached, and after a 2-min rest, the subjects performed a 4-min walking trial at their PWS for data collection. Then, to test the effects of a different walking speed on the results, the younger subjects performed another 4-min walking trial at 120% of PWS. The analysis with 120% of PWS and their corresponding results are presented in Supplementary File 2.

In biomechanics, there is a consensus that the individual PWS is the most comfortable, safest, stable, and balanced walking speed for each subject, making the results more comparable between different groups [37–39].

2.4. Data analysis

The steps were detected as the zero-cross of the heel-marker velocity [40,41]. Then, the 150 intermediate strides were identified, discarding the initial and final strides exceeding 150. Mediolateral (ML), anteroposterior (AP), and vertical (V) T1 marker velocities—obtained from the T1 marker raw data using the three points derivative method [33,40]—were used in this study due to their stationarity, which is a necessary condition for the majority of nonlinear dynamical quantifiers [9,42]. All three directions were evaluated because they can lead to different conclusions depending on the protocol and extracted quantifiers [2,5,43].

All quantifiers were initially calculated for 150 strides, which is the reported minimum number of strides needed for reliable results of some nonlinear descriptors, including RQA quantifiers [17–20,27]. This number of strides is the greatest value found in studies in which RQA was applied to gait data [2,4,9,27].

To test the effect of reducing the number of strides, T1 marker velocity time series were used to extract RQA quantifiers with a length corresponding to 150, 135, 120, 105, 90, 75, 60, 45, 30, 15, and 7 strides. These represent the range from the maximum to the minimum number of reliable strides found in the literature [27]. The number of strides, rather than the length of the time series, was considered because there is a natural fluctuation in stride duration and length within and between subjects.

In order to estimate the lower sampling frequency, the power spectral density of the data was estimated using Welch's method (Fig. 1). As most of the signal power was below 10 Hz, the data were downsampled to 50 Hz and 25 Hz by decimation using the Matlab (version 2017a, MathWorks Inc., Natick, MA) *decimate* function to test the effect of reducing the sampling frequency. By default, the *decimate* function applies an anti-aliasing, 8th order, zero-lag, low pass Chebyshev Type I filter with a normalized cut-off frequency of $0.8/df$, where df is the decimation factor.

Next, a high-dimensional state space was reconstructed from the T1 marker velocity in the ML, AP, and V directions through the delay embedding method [44]. The ideal time delay and embedding dimensions were determined using average mutual information and false nearest neighbor algorithms, respectively [19], for each subject in each condition. The embedding dimensions ranged from 3 to 5, and the time delays from 15 to 37 samples. All data were processed in Matlab using a custom-written code.

2.5. RQA quantifiers and parameterization

A detailed description of the quantifiers extracted from the RP can be found elsewhere [25,26]. Briefly, in the high-dimensional state space, for a given trajectory x_i ($i = 1, 2, \ldots, N, x \in \mathbb{R}^m$), where N is the trajectory length, the recurrences are defined by $R_{i,j}(r) = \Theta(r - ||x_i - x_j||)$, where r is the neighborhood threshold, $||.||$ is the Euclidian norm, and $\Theta(x)$ is the Heaviside function [26]. Next, the RP is constructed; RP is a graphical representation of the recurrences, which are indicated as black points [26]. Then, the quantifiers are extracted from the RP.

The RR (Eq. (1)) is the proportion between the number of recurrent points on the RP and the total number of points in the state space, excluding the diagonal line [25]:

$$RR = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{i,j}. \tag{1}$$

The RR is associated with a measure of the nonlinear autocorrelation of points in the state space and the dynamical system stability [2,30].

The DET (Eq. (2)) is the percentage of points in the state space that forms diagonal lines in the RP with a minimum length of l_{min} points, where l_{min} (tolerance level) is a threshold determined to avoid saturation of the quantifiers excluding the diagonal lines formed by the tangential motion of the phase space trajectory [25,26]. It is found using

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{i,j}^{N} R_{i,j}}, \tag{2}$$

where $P(l)$ is the histogram of diagonal lines of length l. The DET is widely applied for characterizing predictability as well as chaotic and stochastic behavior of a signal, which are related to adaptability [14,25].

The ENT (Eq. (3)) is the Shannon entropy of the diagonal line length distribution. It is related to deterministic structures on the RP, reflects the complexity of the RP with respect to the diagonal lines [26], and is associated with the DET quantifier, which gives a measure of adaptability [14,25]. The ENT is calculated using

$$ENT = -\sum_{l=l_{min}}^{N} p(l)\ln p(l) \tag{3}$$

where $p(l)$ is the probability of finding a diagonal line of exact length l in the RP given by $p(l) = P(l)/N_l$, and $N_l = \sum_{l \geq l_{min}} P(l)$ is the total number of diagonal lines.

The AVG (Eq. (4)) is the average length of diagonal lines in the RP:

$$AVG = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=l_{min}}^{N} P(l)} E \tag{4}$$

The AVG is associated with both DET and ENT. AVG is the average time that two segments of an orbit are close to each other, thus being interpreted as the mean prediction time, the divergence of orbit segments, and stability [26,27].

The $MaxL$ (Eq. (5)) is the maximum length of diagonal lines in the

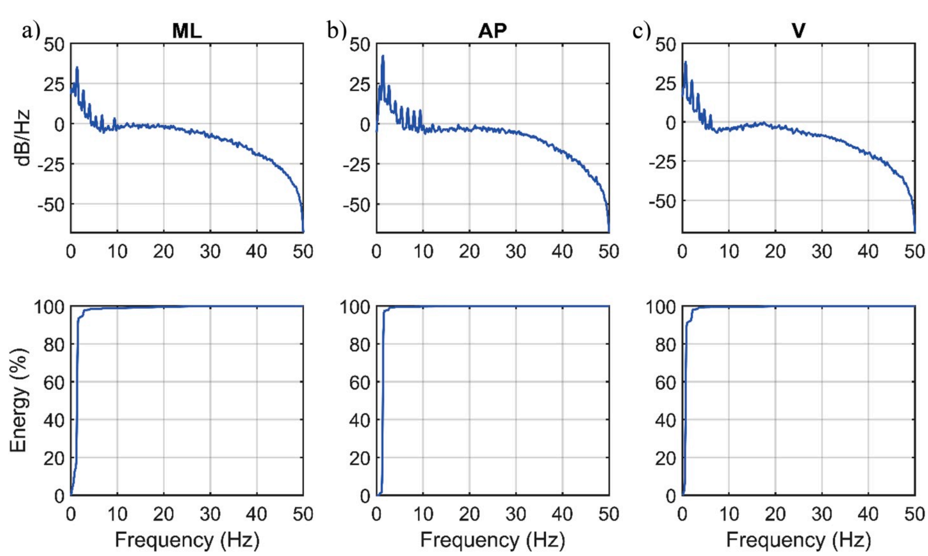

Fig. 1. Welch's power spectral density estimates of (a) mediolateral, (b) anteroposterior, and (c) vertical T1 marker velocity time series. In all directions, most of the signal power is below 10 Hz.

RP:

$$MaxL = \max(\{l_i\}_{i=1}^{N_l}). \tag{5}$$

The *MaxL* is related to the concept of Lyapunov stability, which is the exponential divergence of an orbit [2,26,27].

The *TREND* (Eq. (6)) provides a linear regression coefficient over the recurrence points density that provides information about the stationarity in the process [25]. It is found using

$$TREND = \frac{\sum_{\tau=1}^{\tilde{N}}(\tau - \frac{\tilde{N}}{2})(RR_\tau - RR_\tau)}{\sum_{\tau=1}^{\tilde{N}}(\tau - \frac{\tilde{N}}{2})^2}, \tag{6}$$

where τ is the time distance from the line of identity of a certain line parallel to the line of identity, RR_τ is the corresponding recurrence point density, and $\tilde{N} < N$ excludes the edges of the RP [45].

RQA requires some input parameters to be selected [26]: the threshold r for determining the recurrence points and the minimal length l_{min} of diagonal lines. The adopted threshold r was 40% of the maximum distance between the points in the state space, making the quantifiers behave smoothly and making the results comparable to previous studies [5,15,27]. In addition, l_{min} (for *DET* calculation) was set to 2 so that the diagonals represented determinism [46].

Furthermore, as the results can be sensitive to these input parameters, RQA was conducted using r set to 10% of the maximum distance between the points in the state space [9–12,28], and l_{min} was set to 3 and 5 [9,14]. These values for l_{min} were chosen considering the time interval between two points at 25 Hz (0.04 s), which contain 3 points at 50 Hz (time interval = 0.02 s) and 5 points at 100 Hz (time interval = 0.01 s). The results are presented in Supplementary File 3.

2.6. Statistical analysis

The data normality hypothesis was rejected (Shapiro–Wilk test; $p < 0.05$) for all quantifiers. In order to verify how the quantifiers were altered, two statistical methods were applied. The first was verifying the maximum reduction in the number of strides and in the sampling frequencies that did not alter the median of the values of the quantifiers. The second was verifying the maximum reduction of the parameters studied in order to not deteriorate the within-group relationship (correlation), as shown in Fig. 2.

In order to verify the effects of sampling frequency and the reduction in the number of strides on the RQA quantifiers, we applied the Friedman two-way test ($\alpha = 0.05$) to compare the values of each quantifier in each condition [47]. When the main effect with respect to any of the two factors was significant, a *post-hoc* Wilcoxon signed-rank test was conducted. As the main concern is guaranteeing that the quantifier values are equal when comparing different conditions, no correction (e.g., Bonferroni) was applied to the *post-hoc* test to reduce Type II errors. It was more important to not assume a false reduction [47] because the objective was to find the smallest reduction that does not contain significant differences (i.e., to not reject the null hypothesis). The RQA quantifiers obtained using a sampling frequency of 100 Hz and 150 strides were taken as reference conditions for the Wilcoxon test.

Next, in order to verify if the quantifiers preserved their qualitative characteristics (proportion within individuals) with the reduction in sampling frequency and number of strides, we applied Spearman's correlation using the same reference conditions. The correlation coefficients (ρ) that fell within the intervals $\rho \geq 0.9$, $0.7 < \rho < 0.9$, $0.5 \leq \rho \leq 0.7$, and $\rho < 0.5$ were considered to represent very strong, strong, moderate, and weak correlations, respectively. Only significant correlations ($p < 0.05$) were considered [48]. The proportion among individuals (i.e., the relation of the quantifier's values among individuals in each condition) was considered to be preserved only when Spearman's correlation was very strong.

A condition was considered appropriate compared to the reference condition if it satisfied both Spearman's correlation (significant and very strong) and the Wilcoxon ($p > 0.05$) tests. Statistical analysis was performed using Matlab Statistical Toolbox version 2017a with a significance level set at $\alpha < 0.05$.

3. Results

The RR, DET, AVG, EntD, and TREND quantifiers exhibited similar behavior when increasing the number of strides and sampling frequency in all directions. All values reached a plateau at around 60 strides (Fig. 3).

Except for AVG, the number of strides had a significant main effect on the quantifiers (Friedman test, $p < 0.05$). Similarly, the sampling frequency had a significant main effect on all quantifiers (Friedman test,

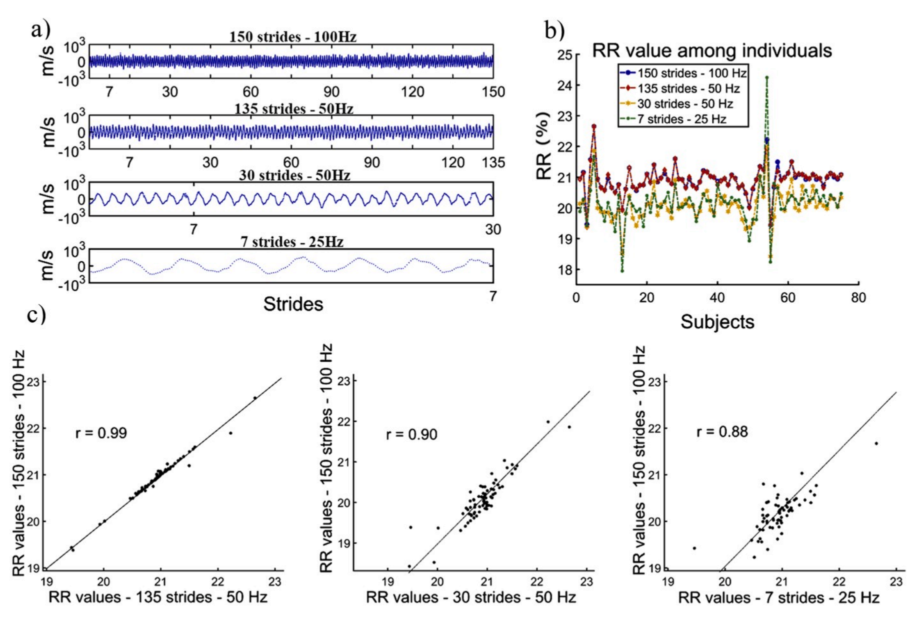

Fig. 2. (a) Typical time series in the ML direction; from top to bottom: 150 strides at 100 Hz, 135 strides at 50 Hz, 30 strides at 50 Hz, and 7 strides at 25 Hz. (b) Plot of the RR quantifier values extracted from the corresponding time series shown in A for each subject. (c) Intra-group deterioration of RR values presented by correlation between reference condition (150 strides at 100 Hz) and other conditions: 135 strides at 50 Hz, 30 strides at 50 Hz, and 7 strides at 25 Hz.

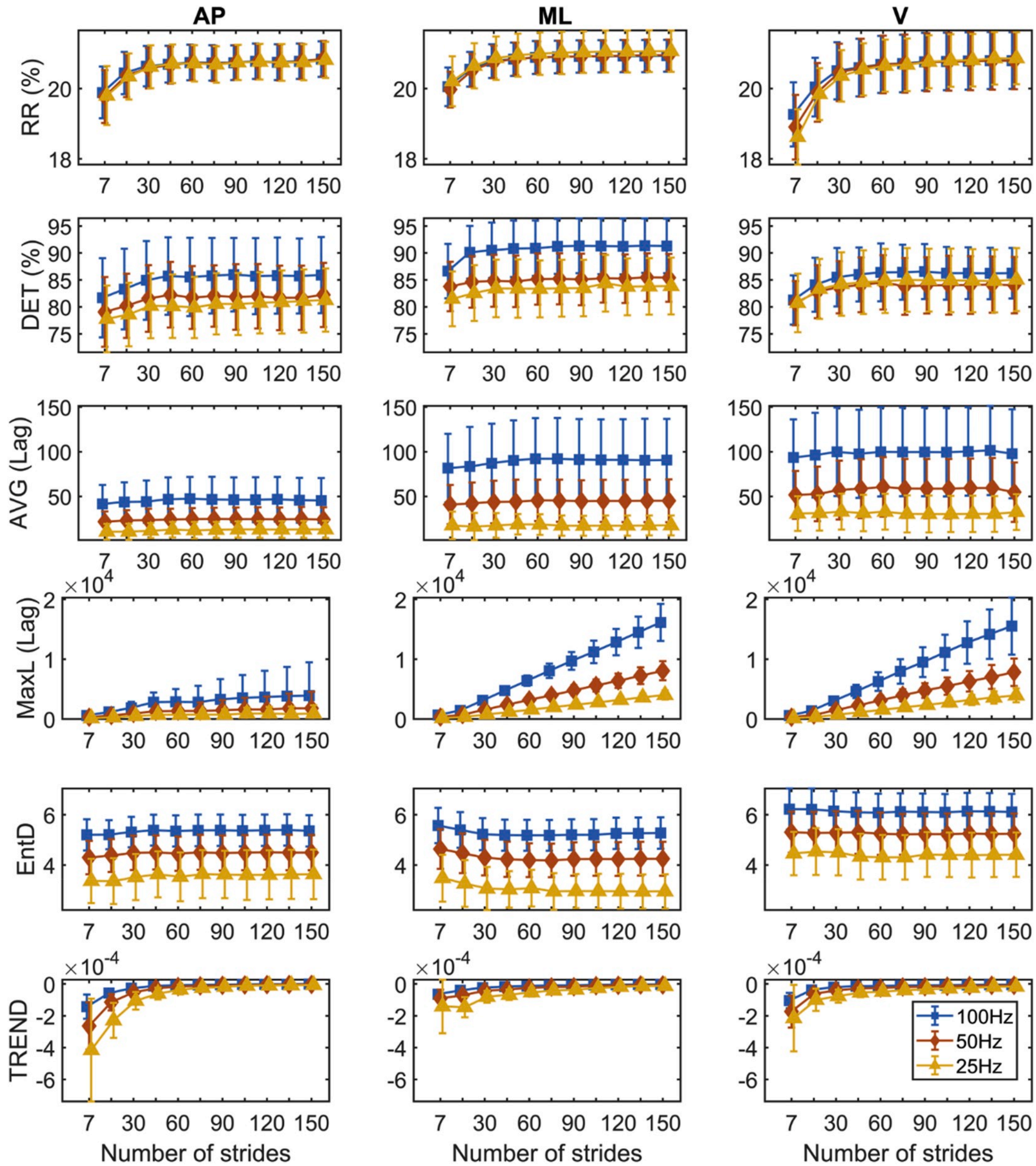

Fig. 3. Medians and standard deviations of the quantifiers in each condition (number of strides and sampling frequency) in the anterior-posterior, medial-lateral, and vertical directions. The threshold r was set to 40% of the maximum distance in the state space, and $l_{min} = 2$.

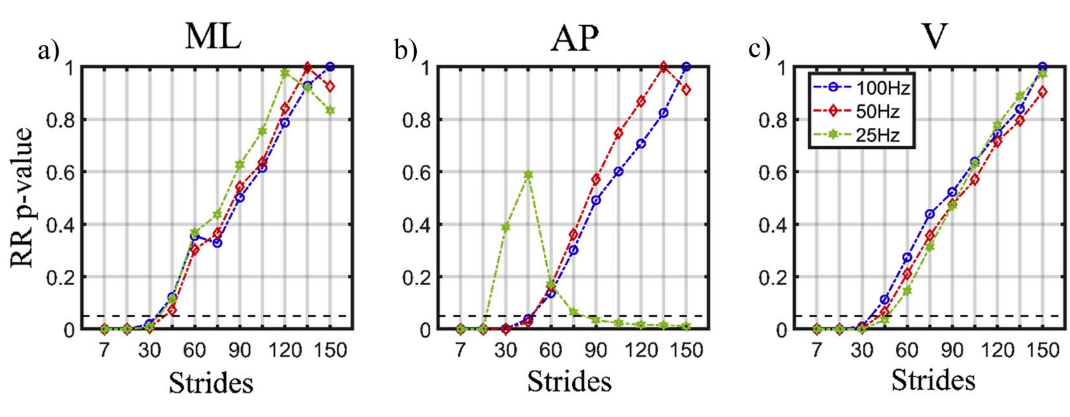

Fig. 4. Significance of the Wilcoxon test for the RR quantifier between each condition and reference condition (100 Hz and 150 strides) in the (a) ML, (b) AP, and (c) V directions. Each plot presents a horizontal dashed line at $p = 0.05$, corresponding to the reduction threshold in the number of strides and sampling frequency. Below this line, the correlation between each condition and control condition was non-significant.

$p < 0.05$).

The DET, AVG, MaxL, and EntD quantifiers were very sensitive to the sampling frequency (Fig. 3). DET, however, exhibited an inverse relationship with sampling frequency: the values of DET increased when the sampling frequency decreased, suggesting that the determinism of a time series does not depend on high-frequency components in the time series. RR did not present an expressive dependence on sampling frequency.

AVG and EntD did not present an expressive dependence on the number of strides. RR, DET, and TREND presented a dependence on the number of strides only for 60 strides or less. In contrast, MaxL was very sensitive, presenting a linear relationship with the number of strides.

However, to propose a reduction in the number of strides and in the sampling frequency, further analysis was needed. Fig. 4 shows the significance (p value) of Spearman's correlation between each condition and the control for the RR quantifier.

According to Fig. 4, the number of strides had a significant main effect on the RR in all directions (ML, AP, and V), and the greatest differences appeared when the number of strides was 45 or less in the ML and AP directions ($p < 0.05$) (Fig. 4a and b, respectively) and 30 in the V direction (Fig. 4c). Thus, the RR quantifier requires a time series with a minimum of 60 strides sampled at 50 Hz in all directions. The Spearman correlation coefficients were very strong for all directions and conditions, down to 15 strides and 25 Hz, as shown in Fig. 5. The same analysis for the other quantifiers is presented in Figs. 6 and 7.

Table 1 summarizes the results for all quantifiers, considering the highest threshold among all directions. The minimum values for sampling frequency and number of strides needed in order to not corrupt the distributions (no significant difference when applying the Wilcoxon signed-rank test) and correlations (very strong and significant) are presented.

4. Discussion

In this study, we analyzed the effect of different sampling frequencies and the number of strides on the RQA quantifiers extracted from the gait time series. The results demonstrated that all quantifiers suffered deterioration to some extent, and the way the different quantifiers were calculated produced distinct effects regarding the number of strides and sampling frequency reduction, as initially hypothesized. As expected, RR, DET, AVG, EntD, and TREND were insensitive to a reduction in the number of strides, at least until ~60 strides, presenting no significant differences when comparing the different conditions to the control condition (Fig. 6) and correlations greater than 0.9 (Fig. 7). However, DET, AVG, and EntD were sensitive to the reduction in sampling frequency, presenting correlations less than 0.9 for 25 Hz. TREND presented no significant differences when comparing the different conditions to the control condition (Fig. 7), but correlations were less than 0.9 below 135 strides (Fig. 7). In contrast, MaxL was very sensitive to both the number of strides and sampling frequency (Fig. 3), presenting correlations less than 0.9 below 105 strides at 50 Hz and 100 Hz and for all numbers of strides at 25 Hz.

Therefore, it is important to take into consideration the number of strides and the sampling frequency for RQA quantifiers extracted from gait kinematics data. Furthermore, in support of the hypothesis, the results enable the proposal of baseline values (Table 1) for the number of strides and sampling frequency in order to preserve the quantifier values' distribution, namely location (median) and scale (standard deviation), as well as the proportion within individuals.

Although the power spectral density analysis using the Welch method showed that kinematic gait data has most of its power content below 10 Hz, the results in Table 1 indicate that the minimum sampling frequency to be adopted for RQA analysis is 100 Hz when using the selected quantifiers simultaneously. This result is in agreement with [49]; who showed that RQA quantifiers are more sensitive to sampling frequency than spectral analysis.

To preserve all characteristics, it is necessary to take the largest result between the distribution and correlation presented in Table 1. For example, the stability and trend measures (MaxL and TREND, respectively) did not allow for any reduction in the number of strides and sampling frequency. This is similar to the results found by Ref. [50] for the Lyapunov exponent and maximum Floquet quantifier, for which they recommended a minimum of 150 strides to return reliable values.

The remaining quantifiers (RR, DET, AVG, and EntD), when analyzed altogether, behaved consistently even with a 60% reduction in the number of strides, corresponding to a final number of 60 strides. However, the sampling frequency must still be maintained at 100 Hz. Thus, the data collection time decreased to less than 2 min and the time series length to around 9600 samples. This can be explained by the fact that these quantifiers are calculated from rates and proportions [25].

The RR, DET, AVG, and EntD quantifiers presented the greatest potential for a reduction in the number of strides, suggesting that less strides can be appropriate when evaluating gait events with short temporal localization on RP (i.e., using small windows on RP). Examples of such events are studies on the reaction time for gait stabilization after perturbations [51], studies concerning resistance/fatiguing protocols [34,49], and studies on gait adaptability under/not under dual-task conditions [52] that would require applying the sliding window method to RP. The last method consists of defining a small window that will slide on the RP line of identity [25]. Thus, the RQA quantifiers will be calculated based on a reduced number of strides, which are contained in this small window.

Overall, though the RQA quantifier values were different, the RQA quantifier's behavior with respect to sampling frequency and number of

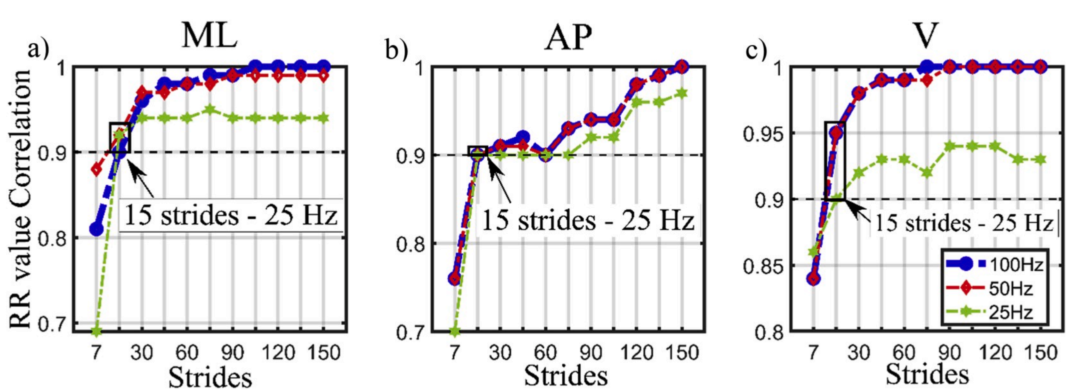

Fig. 5. Correlation coefficient between each condition and the reference condition (100 Hz and 150 strides) for the RR quantifier in the (a) ML, (b) AP, and (c) V directions. The horizontal dashed line indicates the reduction threshold in the number of strides and sampling frequency, above which the correlation was very strong ($\rho \geq 0.9$).

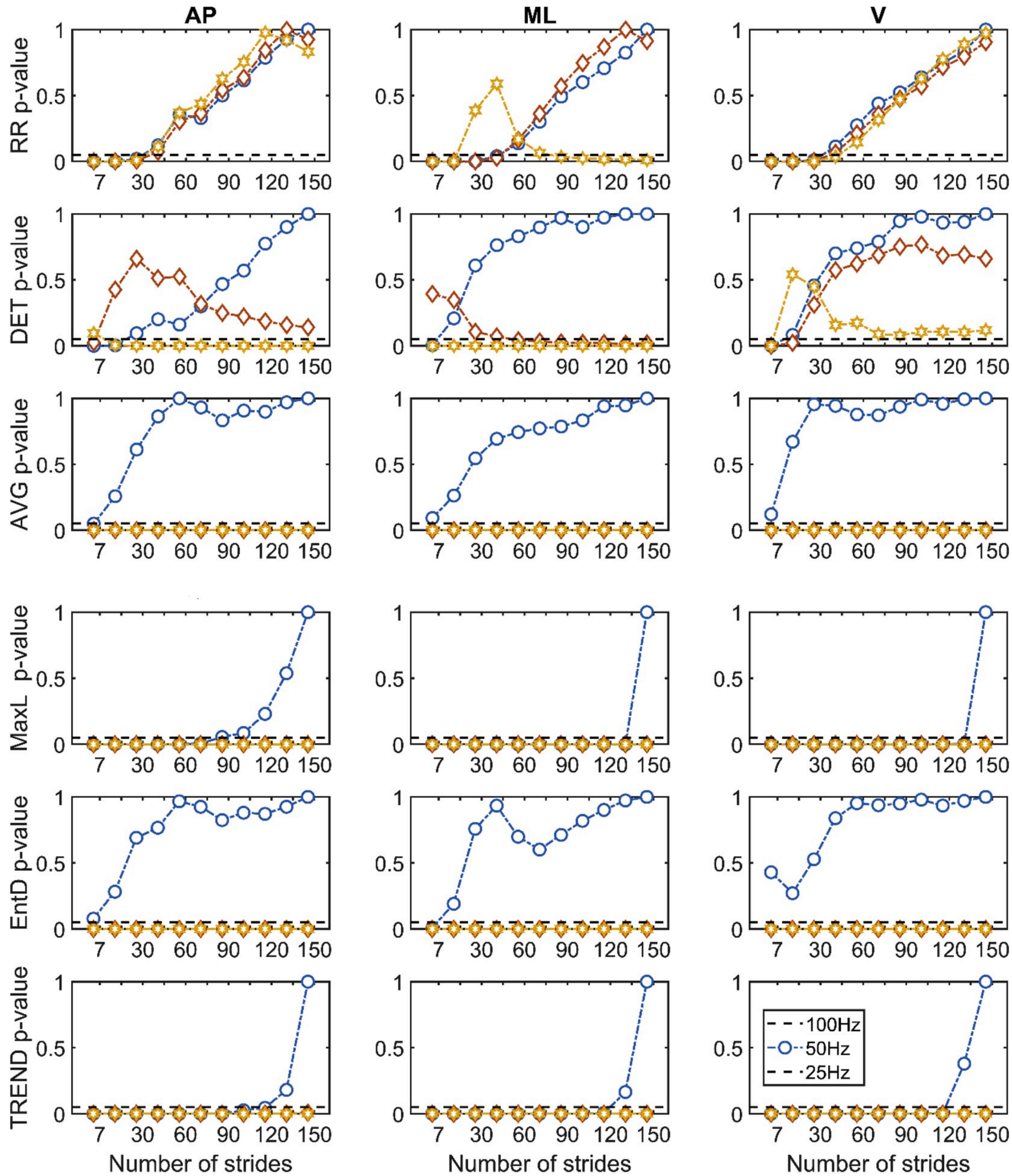

Fig. 6. The *p*-values of the Wilcoxon test applied between the reference condition and the remaining conditions in the anterior-posterior, medial-lateral, and vertical directions for each RQA quantifier. The black horizontal dashed line is the threshold for reduction in the number of strides and sampling frequency, above which the compared quantifier's results were not different. The threshold r was set to 40% of the maximum distance in the state space, and $l_{min} = 2$.

strides was not sensitive to a different walking speed (see Supplementary File 2). In addition, comparing younger and older adults (Supplementary File 1) both walking at their corresponding PWS, the RQA quantifier values were different, but the RQA quantifier's behavior regarding sampling frequency and number of strides remained unchanged. These latter results support the idea that when comparing different populations, these populations should be tested at their corresponding PWS [53,54]. Furthermore, previous studies have shown that RQA quantifiers extracted from treadmill gait data have a good correlation with fall history [55], suggesting a possible clinical application for a more valid and robust fall risk index [27]. Therefore, a reduction in data collection time for extraction of RQA quantifiers would be welcome for populations at high risk of falls. In addition to reducing the computational cost of calculating RQA quantifiers, when dealing with older adults, impaired people, or those suffering from a neuromotor disorder, a reduced data collection time would provide a more comfortable data collection session for the volunteers. However, outcomes with older adults and people with disabilities need to be further confirmed with a larger sample, and those with neuromotor disorders need to be tested.

This study has some limitations. A critical input parameter to the RP is the threshold *r*. Several criteria for the choice of *r* have been proposed in the literature [26]. In the present study, a fixed threshold corresponding to a percentage of the maximum distance between the points in the state space, which is most commonly used in gait studies, was adopted. Thus, RP constructed using other criteria for the choice of *r* should be tested. Another limitation was the necessity of using the same number of strides for all subjects, producing time series with different numbers of points due to small differences in gait speed between the subjects. As shown, some RQA quantifiers are sensitive to data length. This issue was minimized using a sample that was as homogeneous as possible and applying statistical methods in which a subject is compared with himself/herself. However, when comparing populations or

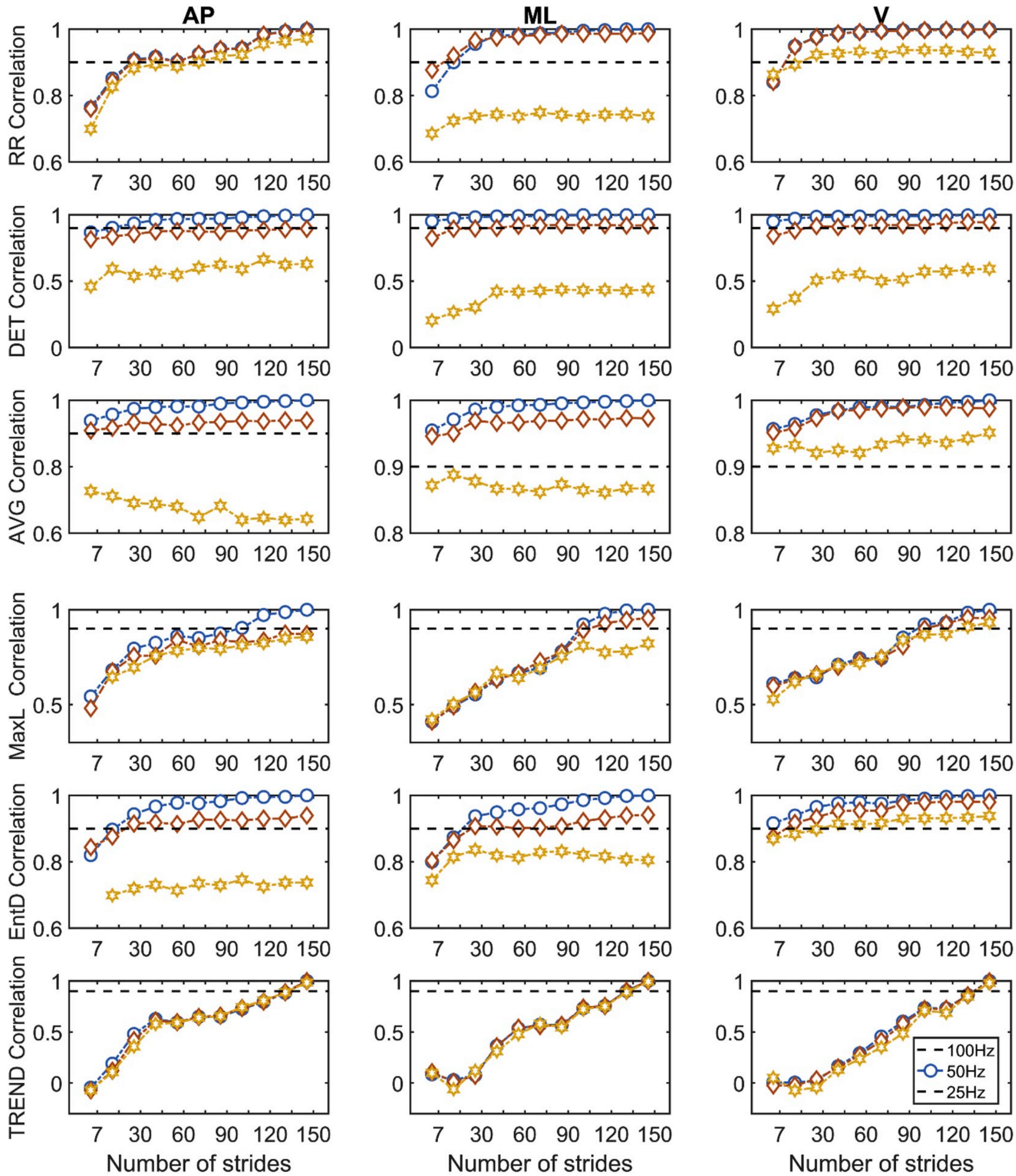

Fig. 7. Spearman correlation for RQA quantifiers between the reference condition and the remaining conditions in the anterior-posterior, medial-lateral, and vertical directions. The black horizontal dashed line is the threshold for reduction in the number of strides and sampling frequency, above which the correlation is very strong. The threshold r was set to 40% of the maximum distance in the state space, and $l_{min} = 2$.

Table 1
Minimum values for the sampling frequency and number of strides needed to preserve the value distributions and correlations of the analyzed RQA quantifiers.

Quantifiers	Distributions		Correlations	
	Strides	SF (Hz)	Strides	SF (Hz)
RR	60	50	15	25
DET	45	100	15	100
AVG	7	100	7	50
MaxL	150	100	105	100
ENT	15	100	30	50
TREND	135	100	150	25

SF = sampling frequency.

situations with considerable differences in gait speed, caution should be used regarding the number of strides: a normalization with a fixed number of strides and a fixed number of samples across all data should be adopted [56,57].

In summary, to simultaneously extract the RR, DET, AVG, and ENT quantifiers while still preserving all properties, a minimum of 60 strides with a 100-Hz sampling frequency is necessary. For MaxL and TREND quantifiers, however, a minimum of 150 strides collected at 100 Hz is recommended. Therefore, caution should be used when comparing studies with a different number of strides and sampling frequency. Furthermore, as shown in Supplementary File 3, RQA quantifiers were sensitive to the parameters r and l_{min}, although the qualitative behavior of such quantifiers (in terms of the number of strides and the sampling frequency) remained unchanged. Thus, caution should also be used when comparing studies with different r and l_{min} values. Further studies should focus on the functional significance of RQA quantifiers computed with different values of r and l_{min}.

5. Conclusion

The main finding of this study is the possibility of a shorter data collection period in terms of gait, with a 60% reduction in the number of strides. However, the sampling frequency must be maintained at 100 Hz for the RR, DET, AVG, and EntD quantifiers to prevent them from losing their statistical characteristics. This allows for the possibility of studies applying the sliding window method to the RP since the correlations within the same subject are preserved with a smaller number of strides.

Declaration of competing interest

None of the authors have conflict of interest.
"None declared".

Acknowledgments

This study was supported by governmental agencies CAPES, CNPq, FAPEMIG, and FAPEG. Adriano O. Andrade and Marcus Fraga Vieira are a Fellow of CNPq, Brazil (304818/2018-6, 306205/2017-3, respectively).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2020.103673.

References

[1] D. Winter, Human balance and posture control during standing and walking, Gait Posture 3 (1995) 193–214, https://doi.org/10.1016/0966-6362(96)82849-9.
[2] M.C. Bisi, F. Riva, R. Stagni, Measures of gait stability: performance on adults and toddlers at the beginning of independent walking, J. NeuroEng. Rehabil. 11 (2014), https://doi.org/10.1186/1743-0003-11-131.
[3] W. Müller, A. Jung, H. Ahammer, Advantages and problems of nonlinear methods applied to analyze physiological time signals: human balance control as an example, Sci. Rep. 7 (2017) 2464, https://doi.org/10.1038/s41598-017-02665-5.
[4] O. Afsar, U. Tirnakli, J. Kurths, Entropy-based complexity measures for gait data of patients with Parkinson's disease, Chaos Interdiscip. J. Nonlinear Sci. 26 (2016), https://doi.org/10.1063/1.4942352, 023115.
[5] L. Bizovska, Z. Svoboda, N. Vuillerme, M. Janura, Multiscale and Shannon entropies during gait as fall risk predictors—a prospective study, Gait Posture 52 (2017) 5–10, https://doi.org/10.1016/j.gaitpost.2016.11.009.
[6] H. Cao, L. Peyrodie, O. Agnani, F. Cavillon, P. Hautecoeur, C. Donzé, Evaluation of an expanded disability status scale (EDSS) modeling strategy in multiple sclerosis, Med. Biol. Eng. Comput. 53 (2015) 1141–1151, https://doi.org/10.1007/s11517-015-1383-7.
[7] D. Hamacher, N.B. Singh, J.H. Van Dieen, M.O. Heller, W.R. Taylor, Kinematic measures for assessing gait stability in elderly individuals: a systematic review, J. R. Soc. Interface 8 (2011) 1682–1698, https://doi.org/10.1098/rsif.2011.0416.
[8] Daniel Hamacher, Dennis Hamacher, R. Müller, L. Schega, A. Zech, Exploring phase dependent functional gait variability, Hum. Mov. Sci. 52 (2017) 191–196, https://doi.org/10.1016/j.humov.2017.02.006.
[9] F.S. Labini, A. Meli, Y.P. Ivanenko, D. Tufarelli, Recurrence quantification analysis of gait in normal and hypovestibular subjects, Gait Posture 35 (2012) 48–55, https://doi.org/10.1016/j.gaitpost.2011.08.004.
[10] M. Mazaheri, H. Negahban, M. Salavati, M.A. Sanjari, M. Parnianpour, Reliability of recurrence quantification analysis measures of the center of pressure during standing in individuals with musculoskeletal disorders, Med. Eng. Phys. 32 (2010) 808–812, https://doi.org/10.1016/j.medengphy.2010.04.019.
[11] H. Negahban, M.A. Sanjari, R. Mofateh, M. Parnianpour, Nonlinear dynamical structure of sway path during standing in patients with multiple sclerosis and in healthy controls is affected by changes in sensory input and cognitive load, Neurosci. Lett. 553 (2013) 126–131, https://doi.org/10.1016/j.neulet.2013.08.024.
[12] H. Negahban, M. Salavati, M. Mazaheri, M.A. Sanjari, M.R. Hadian, M. Parnianpour, Non-linear dynamical features of center of pressure extracted by recurrence quantification analysis in people with unilateral anterior cruciate ligament injury, Gait Posture 31 (2010) 450–455, https://doi.org/10.1016/j.gaitpost.2010.01.020.
[13] P. Prabhu, A.K. Karunakar, H. Anitha, N. Pradhan, Classification of gait signals into different neurodegenerative diseases using statistical analysis and recurrence quantification analysis, Pattern Recogn. Lett. (2018), https://doi.org/10.1016/j.patrec.2018.05.006.
[14] S. Ramdani, G. Tallon, P.L. Bernard, H. Blain, Recurrence quantification analysis of human postural fluctuations in older fallers and non-fallers, Ann. Biomed. Eng. 41 (2013) 1713–1725, https://doi.org/10.1007/s10439-013-0790-x.
[15] P. Tamburini, F. Storm, C. Buckley, M.C. Bisi, R. Stagni, C. Mazzà, Moving from laboratory to real life conditions: influence on the assessment of variability and stability of gait, Gait Posture 59 (2018) 248–252, https://doi.org/10.1016/j.gaitpost.2017.10.024.
[16] S.M. Bruijn, O.G. Meijer, P.J. Beek, J.H. Van Dieen, Assessing the stability of human locomotion: a review of current measures, J. R. Soc. Interface 10 (2013) 20120999, https://doi.org/10.1098/rsif.2012.0999.
[17] S.M. Bruijn, W.R.T. Ten Kate, G.S. Faber, O.G. Meijer, P.J. Beek, J.H. Van Dieën, Estimating dynamic gait stability using data from non-aligned inertial sensors, Ann. Biomed. Eng. 38 (2010) 2588–2593, https://doi.org/10.1007/s10439-010-0018-2.
[18] J.B. Dingwell, J.P. Cusumano, Nonlinear time series analysis of normal and pathological human walking, Chaos 10 (2000) 848–863, https://doi.org/10.1063/1.1324008.
[19] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, 2003.
[20] S.J. McGregor, M.A. Busa, J. Skufca, J.A. Yaggie, E.M. Bollt, Control entropy identifies differential changes in complexity of walking and running gait patterns with increasing speed in highly trained runners, Chaos 19 (2009), https://doi.org/10.1063/1.3147423.
[21] A. Sehle, A. Mündermann, K. Starrost, A. Mundermann, K. Starrost, S. Sailer, I. Becher, C. Dettmers, M. Vieten, Objective assessment of motor fatigue in Multiple Sclerosis using kinematic gait analysis: a pilot study, J. NeuroEng. Rehabil. 8 (2011) 59, https://doi.org/10.1186/1743-0003-8-59.
[22] E. Wass, N.F. Taylor, A. Matsas, Familiarisation to treadmill walking in unimpaired older people, Gait Posture 21 (2005) 72–79, https://doi.org/10.1016/j.gaitpost.2004.01.003.
[23] C.J. Hasson, R.E.A. Van Emmerik, G.E. Caldwell, J.M. Haddad, J.L. Gagnon, J. Hamill, Influence of embedding parameters and noise in center of pressure recurrence quantification analysis, Gait Posture 27 (2008) 416–422, https://doi.org/10.1016/j.gaitpost.2007.05.010.
[24] J.M. Schmit, M.A. Riley, A. Dalvi, A. Sahay, P.K. Shear, K.D. Shockley, R.Y.K. Pun, Deterministic center of pressure patterns characterize postural instability in Parkinson's disease, Exp. Brain Res. 168 (2006) 357–367, https://doi.org/10.1007/s00221-005-0094-y.
[25] C.L. Webber Jr., N. Marwan, Recurrence Quantification Analysis -Theory and Best Practices, Springer Complexy, 2015.
[26] N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, Phys. Rep. (2007), https://doi.org/10.1016/j.physrep.2006.11.001.
[27] F. Riva, M.C. Bisi, R. Stagni, Gait variability and stability measures: minimum number of strides and within-session reliability, Comput. Biol. Med. 50 (2014) 9–13, https://doi.org/10.1016/j.compbiomed.2014.04.001.
[28] C.K. Rhea, T.A. Silver, S.L. Hong, J.H. Ryu, B.E. Studenka, C.M.L. Hughes, J. M. Haddad, Noise and complexity in human postural control: interpreting the different estimations of entropy, PloS One 6 (2011) 1–9, https://doi.org/10.1371/journal.pone.0017696.
[29] R. Müller, P. Büttner, A critical discussion of intraclass correlation coefficients, Stat. Med. (1994), https://doi.org/10.1002/sim.4780132310.
[30] J.M. Schmit, D.I. Regis, M.A. Riley, Dynamic patterns of postural sway in ballet dancers and track athletes, Exp. Brain Res. 163 (2005) 370–378, https://doi.org/10.1007/s00221-004-2185-6.
[31] N. Marwan, How to avoid potential pitfalls in recurrence plot based data analysis, Int. J. Bifurc. Chaos (2011), https://doi.org/10.1142/S0218127411029008.
[32] F.B. Rodrigues, R.M. Magnani, G.C. Lehnen, G.S.D.S.E. Souza, A.O. Andrade, M. F. Vieira, Effects of Backpack Load and Positioning on Nonlinear Gait Features in Young Adults, Ergonomics, 2017, https://doi.org/10.1080/00140139.2017.1413213.
[33] M.F. Vieira, F.B. Rodrigues, G.S. de Sá e Souza, R.M. Magnani, G.C. Lehnen, A. O. Andrade, Linear and nonlinear gait features in older adults walking on inclined surfaces at different speeds, Ann. Biomed. Eng. 45 (2017) 1560–1571, https://doi.org/10.1007/s10439-017-1820-x.
[34] M.F. Vieira, G.S. de Sá e Souza, G.C. Lehnen, F.B. Rodrigues, A.O. Andrade, Effects of general fatigue induced by incremental maximal exercise test on gait stability and variability of healthy young subjects, J. Electromyogr. Kinesiol. 30 (2016) 161–167, https://doi.org/10.1016/j.jelekin.2016.07.007.
[35] C. Mazzà, M. Iosa, F. Pecoraro, A. Cappozzo, Control of the upper body accelerations in young and elderly women during level walking. https://doi.org/10.1186/1743-0003-5-30, 2008.
[36] B.L. Rankin, S.K. Buffo, J.C. Dean, A neuromechanical strategy for mediolateral foot placement in walking humans, J. Neurophysiol. 112 (2014) 374–383, https://doi.org/10.1152/jn.00138.2014.
[37] H.G. Kang, J.B. Dingwell, Separating the effects of age and walking speed on gait variability, Gait Posture 27 (2008) 572–577, https://doi.org/10.1016/j.gaitpost.2007.07.009.
[38] J.B. Dingwell, J.P. Cusumano, D. Sternad, P.R. Cavanagh, Slower speeds in patients with diabetic neuropathy lead to improved local dynamic stability of continuous overground walking, J. Biomech. 33 (2000) 1269–1277, https://doi.org/10.1016/S0021-9290(00)00092-0.
[39] J.B. Dingwell, J.P. Cusumano, Identifying stride-to-stride control strategies in human treadmill walking, PloS One 10 (2015) 1–22, https://doi.org/10.1371/journal.pone.0124879.
[40] G.S. de S. Souza, F.B. Rodrigues, A.O. Andrade, M.F. Vieira, A simple, reliable method to determine the mean gait speed using heel markers on a treadmill, Comput. Methods Biomech. Biomed. Eng. 20 (2017), https://doi.org/10.1080/10255842.2017.1309395.

[41] J.a. Zeni, J.G. Richards, J.S. Higginson, Two simple methods for determining gait events during treadmill and overground walking using kinematic data, Gait Posture 27 (2008) 710–714, https://doi.org/10.1016/j.gaitpost.2007.07.007.
[42] R. Jacob, K.P. Harikrishnan, R. Misra, G. Ambika, Characterization of chaotic attractors under noise: a recurrence network perspective, Commun. Nonlinear Sci. Numer. Simulat. 41 (2016) 32–47, https://doi.org/10.1016/j.cnsns.2016.04.028.
[43] H. Josiński, A. Michalczuk, R. Mucha, A. Świtoński, A. Szczęsna, K. Wojciechowski, Analysis of human motion data using recurrence plots and recurrence quantification measures, 180014, https://doi.org/10.1007/978-3-662-49390-8_39 , 2016, 397-406.
[44] F. Takens, Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics, Springer Berlin Heidelberg, Berlin, Heidelberg, 1981, https://doi.org/10.1007/BFb0091903.
[45] N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, Phys. Rep. 438 (2007) 237–329, https://doi.org/10.1016/j.physrep.2006.11.001.
[46] M.a. Riley, R. Balasubramaniam, M.T. Turvey, Recurrence quantification analysis of postural fluctuations, Gait Posture 9 (1999) 65–78, https://doi.org/10.1016/S0966-6362(98)00044-7.
[47] J.P. Marques De Sá, Applied statistics using SPSS, STATISTICA, MATLAB and R, applied statistics using SPSS, STATISTICA, MATLAB and R. https://doi.org/10.1007/978-3-540-71972-4, 2007.
[48] M. Kraska-MIller, Nonparametric Statistics for Social and Behavioral Sciences, 2013.
[49] C.L. Webber Jr., M.A. Schmidt, J.M. Walsh, Influence of isometric loading on biceps EMG dynamics as assessed by linear and nonlinear tools, J. Appl. Physiol. 78 (1995) 814–822, https://doi.org/10.1152/jappl.1995.78.3.814.
[50] S.M. Bruijn, J.H. van Dieën, O.G. Meijer, P.J. Beek, Statistical precision and sensitivity of measures of dynamic gait stability, J. Neurosci. Methods 178 (2009) 327–333, https://doi.org/10.1016/j.jneumeth.2008.12.015.
[51] M.J.P. Toebes, M.J.M. Hoozemans, J. Dekker, J.H. van Dieën, Effects of unilateral leg muscle fatigue on balance control in perturbed and unperturbed gait in healthy elderly, Gait Posture 40 (2014) 215–219, https://doi.org/10.1016/j.gaitpost.2014.03.194.
[52] L. Hak, H. Houdijk, F. Steenbrink, A. Mert, P. van der Wurff, P.J. Beek, J.H. Van Dieën, Stepping strategies for regulating gait adaptability and stability, J. Biomech. 46 (2013), https://doi.org/10.1016/j.jbiomech.2012.12.017.
[53] C. Moraiti, N. Stergiou, S. Ristanis, A.D. Georgoulis, ACL deficiency affects stride-to-stride variability as measured using nonlinear methodology. Knee Surgery, Sport. Traumatol. Arthrosc. 15 (2007) 1406–1413, https://doi.org/10.1007/s00167-007-0373-1.
[54] S.R. Wurdeman, S.A. Myers, N. Stergiou, Transtibial amputee joint motion has increased attractor divergence during walking compared to non-amputee gait, Ann. Biomed. Eng. 41 (2013) 806–813, https://doi.org/10.1007/s10439-012-0705-2.
[55] F. Riva, M.J.P. Toebes, M. Pijnappels, R. Stagni, J.H. van Dieën, Estimating fall risk with inertial sensors using gait stability measures that do not require step detection, Gait Posture 38 (2013) 170–174, https://doi.org/10.1016/j.gaitpost.2013.05.002.
[56] P.C. Raffalt, J.A. Kent, S.R. Wurdeman, N. Stergiou, Selection procedures for the largest Lyapunov exponent in gait biomechanics, Ann. Biomed. Eng. 47 (2019) 913–923, https://doi.org/10.1007/s10439-019-02216-1.
[57] J. Stenum, S.M. Bruijn, B.R. Jensen, The effect of walking speed on local dynamic stability is sensitive to calculation methods, J. Biomech. 47 (2014) 3776–3779, https://doi.org/10.1016/j.jbiomech.2014.09.020.