INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Hussain VS, Spano ML, Lockhart TE. 2020 Effect of data length on time delay and embedding dimension for calculating the Lyapunov exponent in walking. *J. R. Soc. Interface* **17**: 20200311.
http://dx.doi.org/10.1098/rsif.2020.0311

Received: 30 April 2020
Accepted: 16 June 2020

Subject Category:
Life Sciences–Engineering interface

Subject Areas:
bioengineering, biocomplexity, biomechanics

Keywords:
gait, Lyapunov exponents, local dynamic stability, phase space reconstruction

Authors for correspondence:
Victoria Smith Hussain
e-mail: vasmith5@asu.edu
Thurmon E. Lockhart
e-mail: thurmon.lockhart@asu.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5042045.

THE ROYAL SOCIETY PUBLISHING

Effect of data length on time delay and embedding dimension for calculating the Lyapunov exponent in walking

Victoria Smith Hussain, Mark L. Spano and Thurmon E. Lockhart

School of Biological and Health Systems Engineering, Ira A. Fulton Schools of Engineering, Arizona State University, Tempe, AZ 85287-9709, USA

VSH, 0000-0001-9652-8325; TEL, 0000-0002-7008-5711

The Lyapunov exponent (LyE) is a trending measure for characterizing gait stability. Previous studies have shown that data length has an effect on the resultant LyE, but the origin of why it changes is unknown. This study investigates if data length affects the choice of time delay and embedding dimension when reconstructing the phase space, which is a requirement for calculating the LyE. The effect of three different preprocessing methods on reconstructing the gait attractor was also investigated. Lumbar accelerometer data were collected from 10 healthy subjects walking on a treadmill at their preferred walking speed for 30 min. Our results show that time delay was not sensitive to the amount of data used during calculation. However, the embedding dimension had a minimum data requirement of 200 or 300 gait cycles, depending on the preprocessing method used, to determine the steady-state value of the embedding dimension. This study also found that preprocessing the data using a fixed number of strides or a fixed number of data points had significantly different values for time delay compared to a time series that used a fixed number of normalized gait cycles, which have a fixed number of data points per stride. Thus, comparing LyE values should match the method of calculation using either a fixed number of strides or a fixed number of data points.

1. Introduction

The Lyapunov exponent (LyE) is a nonlinear dynamical calculation that quantifies the rate of divergence or convergence of trajectories in an m-dimensional phase space. A phase space is a finite-dimensional vector space $\mathbb{R}^m$ that contains all the possible states of a system. Each possible state corresponds to one unique point in the phase space and is used to identify the *attractors* in the system. The phase space shows all of the possible trajectories for a dynamical system and is used to identify all of the possible attractors of the system. An attractor draws (repels) nearby trajectories towards (away) from itself, where multiple attractors can combine these properties, repelling in one direction and attracting in another [1–3], also known as saddle points. In classic examples of chaos theory, the phase space is usually a plot of position and momentum as a function of time. However, a phase space can also be reconstructed from a single continuously recorded variable, given that the sampling frequency and the number of cycles of the system are sufficient. Nonlinear dynamics and chaos theory attempt to describe and extract features of these systems to understand their behaviour and sensitivity to initial conditions. For more details and mathematical description of the phase space, readers can refer to [2,4,5]. LyE, or local dynamic stability, is a popular approach to assess and enumerate an individual's ability to withstand minute perturbations during gait. For instance, in walking, we take very similar steps from right to left in terms of step size, walking velocity, etc., but these similar steps are not identical. These small changes are owing to slightly different initial conditions before each step is taken. LyE evaluates these changes (divergences) between initial conditions and is used to measure the stability of gait as a dynamical system. This nonlinear measure has been used to differentiate between healthy and fall prone elderly [6,7], as well as used to identify differences between healthy controls and patients with Parkinson's disease [8], and developmental disorders [9].

Multiple studies have found that the amount of gait data used when calculating the LyE affects the final outcome [10–12]. Previous studies on the reliability of LyE have found different data minimum requirements; some required 54 and 150 strides [13,14], while others state a time duration minimum of 2–3 min of walking data [15,16] is sufficient. However, no studies have investigated if data length plays a role in selecting the reconstruction parameters required for calculating LyE.

The first step in calculating the LyE is reconstructing the collected time series into the phase space so the gait attractor can be analysed. The phase space is reconstructed using the method of delays [17]:

$$y(n) = [x(n), x(n+\tau), \ldots, x(n+(d_E-1)\tau)], \tag{1.1}$$

which requires a time delay, τ, and an embedding dimension, d_E. The time delay is most commonly [18] determined using the first minimum of average mutual information (AMI) function, which evaluates the amount of information that is shared between datasets over a range of time delays. With time delay established, the embedding dimension is then determined using global false nearest neighbours (FNN). FNN compares the distances between neighbouring trajectories at increasing dimensions. False neighbours occur when trajectories overlap in a lower dimension but do not overlap in a larger dimension [19]. The total percentage of false neighbours declines as embedding dimensions increase until the proper embedding dimension is reached. This is usually determined by the FNN percentage as it either approaches zero or plateaus out.

In addition to the varying data lengths being used, previous studies have also applied different preprocessing methods for gait time-series normalization. This has also been found to have an effect on the calculation of LyE [20]. We have identified three major methods in the gait literature:

(1) fixed number of strides with a variable number of total data points [21]—the time series will start and end on a heel contact, but each stride will contain a variable number of data points. This method maintains the distance between points on the attractor;
(2) fixed number of strides and data points per stride [10,11,22]—the time series is time normalized to 100 samples per stride. This method alters the distance between data points within the phase space, but the number of points in each stride cycle is constant across subjects irrespective of gait speed; and
(3) fixed number of data points with a variable number of strides [23,24]—the time series starts at the same as methods 1 and 2 at a heel contact; however, the endpoint is a fixed number of points regardless of the number of gait cycles it contains. This method also maintains the distance between points on the attractor but does not guarantee ending on a full cycle.

The aim of this study was to determine the effect of data length on the reconstruction parameters of the LyE, specifically the τ and d_E determined by AMI and FNN, respectively. We hypothesize that τ and d_E will not change with respect to data length given sufficient data are provided. Additionally, we investigated the effects of three data preprocessing methods on determining time delay and embedding dimension.

2. Material and methods

2.1. Participants

Ten young healthy subjects (five males and five females) with a mean ± standard deviation age of 24.5 ± 4.1 years, body height of 1.67 ± 0.10 m and body mass of 69.4 ± 11.6 kg were included in this study. All subjects were physically active and familiar with walking on a treadmill. Subjects reported no cardiovascular issues, neurological diseases, nor lower extremity surgeries in the last three months. Subjects provided written informed consent before participating in this study. This study was approved by the Institutional Review Board of Arizona State University (ASU FWA 00009102).

2.2. Experimental procedure

After subjects became familiar with the treadmill, each subject's preferred walking speed (PWS) was determined using a standardized protocol [23,25]. The mean and standard deviation of PWS was 1.13 ± 0.1 m s^{-1}. After a short rest period, each subject walked on the treadmill for 30 min at their PWS. Participants wore three tri-axial acceleration sensors sampling at 128 Hz (APDM (Ambulatory Parkinson's Disease Monitoring); Mobility Lab, APDM, Inc., Portland, OR) fitted with elastic bands and Velcro straps and were placed at each ankle and the lower lumbar around vertebrae L4 and L5. For this study, the ankle sensors were used to define heel contacts for truncating the gait data as necessary. A custom algorithm based on previously published algorithms [26,27] was used to define heel contacts. The lumbar sensor was used for reconstructing the phase space and calculating the LyE. The treadmill used in this experiment is a split-belt treadmill and is a part of the GRAIL system (Motekforce Link, Amsterdam, The Netherlands). Measurements were started after the treadmill and the subject was at a constant speed.

Three-dimensional acceleration data of the lumbar sensor were used for all of the calculations in this paper. The heel contacts for each step were determined and indexed, and the time series was truncated to start and end on a heel contact [11,28]. To investigate how different methods of preprocessing affect the calculation of time delay and embedding dimension, three different methods that are used in nonlinear dynamical calculations for gait were implemented:

1. fixed number of strides with a variable number of points per stride;
2. fixed number of strides with 100 data points per stride; and
3. fixed number of data points.

These methods were applied to the vertical (VT), anteroposterior (AP) and mediolateral (ML) acceleration time series and no other filtering/normalization methods were used. After the data were preprocessed, different sample lengths ranging from 30 to 500 strides were extracted from the same first heel contact of the time series. This was repeated for each acceleration direction. The data lengths selected for method 3 were based on 15, 30 and 60 s and 2, 3, 5 and 10 min of gait data. This range includes smaller and larger data collection times as well as very common data collection times of 1–3 min of data. All calculations were done using custom made MATLAB (v. 2018b, Mathworks Inc., Natwick) programs.

2.3. Simulated data

We simulated the Lorenz and Rössler attractors because they are well-known dynamical systems and they are similar to human posture and gait data, respectively. The Lorenz system has a pronounced non-periodic behaviour which may be considered representative for postural sway, while the Rössler system has a periodic behaviour which is more comparable to gait [29]. The systems, based on the differential equations and initial conditions outlined in table 1, were simulated using MATLAB. Each nonlinear attractor was generated with 1×10^6 samples, where the first 8000

Table 1. Reference table for known chaotic dynamical systems. (Values from [30].)

system	equations	parameters	Δt	expected λ_1
Lorenz[a]	$\dot{x} = \sigma(y - x)$	$\sigma = 16.0$	0.01	1.50
	$\dot{y} = x(R - z) - y$	$R = 45.92$		
	$\dot{z} = xy - bz$	$b = 4.0$		
Rössler[b]	$\dot{x} = -y - z$	$a = 0.15$	0.10	0.090
	$\dot{y} = x + ay$	$b = 0.20$		
	$\dot{z} = b + z(x - c)$	$c = 10.0$		

[a]Wolf *et al.* [31].
[b]Rossler [32].

samples were discarded to avoid transient confounders with each time series. Each time series was then segmented into non-overlapping windows that each contained 5×10^4 samples. Ten of these windows were used in the subsequent analyses for both the Lorenz and Rössler attractors. To investigate the effect of data length, various data lengths were extracted from each window ranging from 500 to 5.0×10^4 samples all starting with the first point of the window. This range was used to mimic the data lengths extracted from the gait data using method 3 (data truncated based on a specific number of samples).

2.4. Data analysis

Time delay, τ, was determined as the first local minimum of the AMI function [18]. A time delay was determined for each directional acceleration as data length was varied for the simulated and collected data. The τ determined from AMI at 1×10^4 samples for known systems and 300 gait cycles or 1.5×10^4 data points for gait data. FNN [19,33] was then used to determine the appropriate embedding dimension, d_E, using values of $R_{\mathrm{tol}} = 15$ and $A_{\mathrm{tol}} = 4$. These threshold values within the FNN algorithm are within the suggested ranges set by Kennel *et al.* [19]. The final selection of the d_E is generally up to the discretion of the researcher where the FNN starts plateauing out. Therefore, to objectively select the d_E, we added the following criteria: (i) the difference between subsequent dimensions must be less than 0.05, and (ii) the actual percentage of FNN at that dimension must also be less than 10%. This method is depicted in figure 1. These decision criteria were used for both the Lorenz system and all gait data collected. However, the second criterion had to be increased to 20% for the Rössler system because some subjects in the z-axis, at certain time epochs, never dropped below a 10% FNN rate. This was most likely owing to the fact that less information about the Rössler system lies in the z-axis as most of its' trajectories live in the xy-plane of the phase space.

2.5. Statistical analysis

To explore the effect of data length and preprocessing methods effects on τ and d_E, a repeated measures mixed model ANOVA (analysis of variance) was performed for each signal direction (AP, ML and VT) for the gait data. Data length and preprocessing methods were considered within-subject factors because each subject participated in every data length and all three preprocessing methods. For the simulated nonlinear system data, a one-way repeated measures ANOVA was used to determine the effect of data length on τ and d_E for each signal direction (x, y and z). Subjects were treated as a random factor. Tukey honest significant difference test was used for all post hoc comparisons. For all statistical tests, a

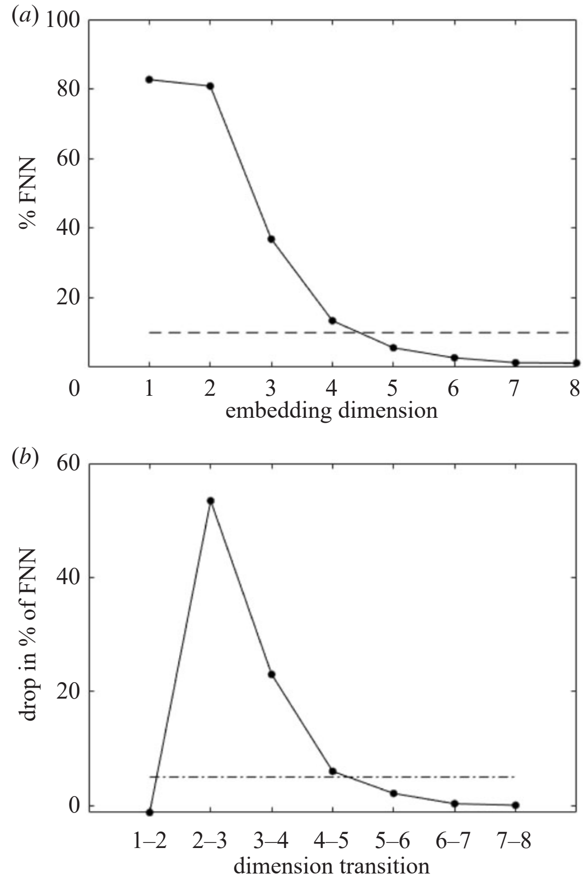

Figure 1. Methodology used to objectively select the embedding dimension. (*a*) The output of FNN. (*b*) The figure was created by finding the difference between neighbouring dimensions, each named for the transition they represent. The first criterion is found when the difference between dimensions is less than 0.05, displayed as the dash-dot line in (*b*). For example, this point would be the 5–6 dimension transition. The second criterion then checks that dimension 5 has less than 10% FNN rate.

p-value < 0.05 was considered significant. All statistical analysis was performed in JMP Pro (v. 14, SAS, Cary, NC).

3. Results

3.1. Simulated systems

There was no statistical effect of data length on τ for the Lorenz attractor in any direction. The Rössler attractor had significant differences in the y-direction ($F_{7,63} = 5.94$, $p = 0.0001$) and the z-direction ($F_{7,63} = 2.81$, $p = 0.0451$). A post hoc Tukey test revealed significant differences between data lengths of 500 points and 2.0×10^3 ($p = 0.0001$), 5.0×10^3 ($p = 0.0007$) and 3.0×10^4 samples ($p = 0.0007$) points, respectively, in the y-direction. In the z-direction, there was only a significant difference between 5.0×10^3 and 1.0×10^4 data points ($p = 0.0451$).

In the Lorenz ($F_{7,63} = 29.22$, $p < 0.0001$; $F_{7,63} = 81.00$, $p < 0.0001$; and $F_{7,63} = 21.00$, $p < 0.0001$ for x, y and z, respectively) and Rössler attractor ($F_{7,63} = 39.86$, $p < 0.0001$; $F_{7,63} = 40.35$, $p < 0.0001$; and $F_{7,63} = 2.81$, $p = 0.0130$ for x, y and z, respectively), data length did have an effect on the d_E. A post hoc Tukey test was used as a pairwise comparison across all data lengths to determine where data length affected d_E. Significant differences ($p = 0.0001$, for all) between data lengths of 500 and any other data length were seen in all directions for the Lorenz system. Additionally, significant differences were

Table 2. The effect of data length when calculating the embedding dimension. (A post hoc Tukey test was used to determine the significant differences between matched data lengths pairwise comparisons. Data length is in gait cycles (gc). Significant differences are in bold.)

		data length (gc)				
	data length (gc)	30	50	100	200	300
AP	30	—	—	—	—	—
	50	0.9684	—	—	—	—
	100	0.3766	0.8438	—	—	—
	200	**<0.0001**	**<0.0001**	**<0.0001**	—	—
	300	**<0.0001**	**<0.0001**	**<0.0001**	**0.0114**	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	**0.0003**	0.8438
ML	30	—	—	—	—	—
	50	1.0000	—	—	—	—
	100	**0.0251**	**0.0251**	—	—	—
	200	**<0.0001**	**<0.0001**	**0.0033**	—	—
	300	**<0.0001**	**<0.0001**	**<0.0001**	0.1380	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	**0.0094**	0.8856
VT	30	—	—	—	—	—
	50	0.9907	—	—	—	—
	100	0.3117	0.6712	—	—	—
	200	**0.0007**	**0.0043**	0.1828	—	—
	300	**0.0001**	**0.0007**	**0.0491**	0.9907	—
	500	**<0.0001**	**<0.0001**	**0.0001**	**0.0984**	0.3117

seen in the x-direction when 2.0×10^3 data points were compared against any other data length ($p = 0.0405$ at 5.0×10^3 data points and $p = 0.0001$ for all other data lengths) in the Lorenz system. In the Rössler system, significant differences ($p = 0.0001$) were seen in the x-direction when 500 and 2.0×10^3 data points were compared against any other data length. In the y-direction, only significant differences ($p = 0.0001$) when 500 data points were compared against any other data length, and no significant differences were found between data lengths in the z-direction. All statistical analysis tables are shown in the electronic supplementary material.

3.2. Gait

Data length did not have an effect on τ in any direction when using gait data. Data length did have an effect on the d_E in all directions (AP: $F_{5,45} = 53.34$, $p < 0.0001$; ML: $F_{5,45} = 45.73$, $p < 0.0001$; and VT: $F_{5,45} = 15.62$, $p < 0.0001$). The post hoc Tukey test revealed that embedding dimension calculations that used very low (30 and 50 gait cycles) or very large (300 and 500 gait cycles) data lengths were, in general, not significantly different from each other in any direction. Table 2 shows all the data length pairwise comparisons.

We found significant differences between the values of τ derived from different preprocessing methods in the AP ($F_{2,18} = 19.61$, $p = 0.001$) and VT directions ($F_{2,18} = 5.05$, $p = 0.0182$). The post hoc Tukey test showed that in these directions, method 1 (fixed number of strides) and method 2 (fixed number of strides with 100 points per stride) were significantly different (AP: $p = 0.0001$ and VT: $p = 0.0413$). Method 2 was also significantly different (AP: $p = 0.0002$ and VT: $p = 0.0275$) from method 3 (fixed number of data points) in the AP and VT directions. These differences are further broken down by data length and are presented in table 3. In the AP direction, all matching data lengths are significantly different when comparing method 1 and method 2. This is also seen when comparing method 2 against method 3, except for when data length is 200 gait cycles. In the VT direction, however, the only significant difference was between method 2 and method 3 when data length was 500 gait cycles. There were significant differences between different methods at different data lengths, but they are not reported in this manuscript as they do not hold relevant information for this analysis.

There were also significant differences between d_E values when calculated with different preprocessing methods, but only in the ML direction ($F_{2,18} = 4.71$, $p = 0.0227$) and specifically only when comparing method 1 to method 2 ($p = 0.0257$). Method differences were further broken down by data length for each signal direction (table 3); there was only a significant difference between method 1 and method 2 when 200 gait cycles were used. We also performed another post hoc Tukey test from a one-way repeated measures ANOVA when data length was a within-subject factor and each preprocessing method was fitted separately. This was done to compare the effect of data length on d_E with respect to each preprocessing method. The results of these analyses can be found in tables 4–6 where bolded values indicate significant differences.

4. Discussion

The time delay and embedding dimension are critical inputs for reconstructing the phase space [34] which is the first step in calculating the LyE. A previous study [14] found that LyE

Table 3. Post hoc Tukey test results for comparing preprocessing methods at each data length for time delay (τ) and embedding dimension (d_E). (Data length is in gait cycles (gc). Significant differences are in bold. Method 1: fixed number of gait cycles. Method 2: fixed number of normalized gait cycles. Method 3: fixed number of data points.)

		time delay			embedding dimension		
	data length (gc)	1 versus 2	1 versus 3	2 versus 3	1 versus 2	1 versus 3	2 versus 3
AP	30	**0.0060**	1.0000	**0.0060**	1.0000	1.0000	1.0000
	50	**0.0153**	1.0000	**0.0113**	0.9600	1.0000	0.9600
	100	**0.0016**	1.0000	**0.0022**	0.4912	1.0000	0.0813
	200	**0.0003**	0.9487	0.1019	1.0000	0.4912	0.4912
	300	**0.0001**	1.0000	**<0.0001**	0.9600	1.0000	1.0000
	500	**0.0011**	1.0000	**0.0016**	1.0000	1.0000	1.0000
	All	**0.0001**	0.9343	**0.0002**	0.2075	0.7323	0.0523
ML	30	0.4120	1.0000	0.4120	1.0000	1.0000	1.0000
	50	0.9697	1.0000	0.9861	1.0000	1.0000	1.0000
	100	0.8421	1.0000	0.8998	0.0813	1.0000	0.0813
	200	0.9697	1.0000	0.9861	**0.0057**	0.4912	0.9600
	300	0.5015	1.0000	0.5015	1.0000	1.0000	1.0000
	500	0.8998	1.0000	0.8998	1.0000	1.0000	1.0000
	All	0.0918	0.9974	0.1042	**0.0257**	0.8609	0.0733
VT	30	0.9760	1.0000	0.9993	1.0000	1.0000	1.0000
	50	0.8325	1.0000	0.5383	1.0000	1.0000	1.0000
	100	0.5383	1.0000	0.5383	0.9447	1.0000	1.0000
	200	0.3369	1.0000	0.5383	0.9447	1.0000	0.9447
	300	0.7456	1.0000	0.7456	0.9447	1.0000	0.4225
	500	0.5383	0.9971	**0.0253**	1.0000	1.0000	1.0000
	All	**0.0413**	0.9784	**0.0275**	0.8601	1.0000	0.8602

increases as data length increases. The specific aspect of the LyE calculation that is sensitive to data length is still unknown. Therefore, this paper investigated the role of data length in the calculation of τ and d_E. Time delay and embedding dimension were calculated using AMI and FNN, respectively. We found that τ is not affected by data length, while d_E is underestimated without sufficient data for its calculation. Additionally, this paper found that stride normalization (method 2) has statistically different τ values compared to gait data that has not been normalized (method 1 and 3). Method 2 generally had smaller τ values in VT and ML directions but had larger values in the AP direction.

As hypothesized, the τ from the Lorenz and walking data does not change as data length increases, regardless of the directional vector. The Rössler system, however, was affected, but only in the y-directional time series. Of the simulated systems used in this study, the Lorenz attractor converged on a time delay of 11 points as reported in a previous study [30]. The time delay of the Rössler attractor was highly variable subject to subject in the x-, y- and z-direction. However, once a data length of 7.5×10^3 points or greater was used, a stable τ (electronic supplementary material, figure S1) was able to be established in all directions.

In gait, no significant differences were found between method 1 (fixed number of strides) and method 3 (fixed number of points) regardless of data length and signal direction. This was expected as the same data were essentially used but truncated differently. Method 2 (fixed number of strides with 100 points per stride) was significantly different from method 1 and method 3, specifically in the AP direction. We found that regardless of data length method 1 and method 3 had significantly larger τ values compared to method 2. It is not surprising that method 2 would have different τ values since every stride is normalized to 100 points per stride. Stride time normalization alters the time and distance relationships within the phase space. It is possible that because the largest movement in walking is moving forward, that the AP direction shows this difference the most. Additionally, the significant differences between these methods in the AP direction may result from position changes, i.e. from the centre to the top of the treadmill or vice versa. Overall, preprocessing methods mainly affected the time delay when calculated from AP data, as shown in figure 2.

Time delay in gait is not as uniform as in the known dynamical systems. The known systems had a single point range about the mean τ, once a sufficient amount of data was used. In gait, the τ ranged from 4 to 16 across all subjects, while the Lorenz and Rössler simulated subjects' time delay ranged from 10 to 12 and 11 to 16, respectively. This larger range is expected owing to the individual gait differences. However, this does beg the question, can the same time delay be used for every subject as well as for each acceleration direction?

Table 4. Post hoc Tukey test results comparing embedding dimensions at all data lengths for the gait data processed using method 1 (fixed number of strides).

		data length comparison for method 1				
	data length (gc)	30	50	100	200	300
AP	30	—	—	—	—	—
	50	1.0000	—	—	—	—
	100	0.9586	0.9586	—	—	—
	200	**<0.0001**	**<0.0001**	**<0.0001**	—	—
	300	**<0.0001**	**<0.0001**	**<0.0001**	0.9586	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	0.1412	0.5536
ML	30	—	—	—	—	—
	50	0.9509	—	—	—	—
	100	**0.0010**	**0.0130**	—	—	—
	200	**<0.0001**	**<0.0001**	**0.0010**	—	—
	300	**<0.0001**	**<0.0001**	**0.0010**	1.0000	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	0.9509	0.9509
VT	30	—	—	—	—	—
	50	1.0000	—	—	—	—
	100	0.3086	0.3086	—	—	—
	200	0.0772	0.0772	0.9812	—	—
	300	**0.0131**	**0.0131**	0.7290	0.9812	—
	500	**0.0002**	**0.0002**	0.0772	0.3086	0.7290

Table 5. Post hoc Tukey test results comparing embedding dimensions at all data lengths for the gait data processed using method 2 (fixed number of strides with 100 points per stride).

		data length comparison for method 2				
	data length (gc)	30	50	100	200	300
AP	30	—	—	—	—	—
	50	0.6756	—	—	—	—
	100	0.0508	0.6756	—	—	—
	200	**<0.0001**	**0.0070**	0.2460	—	—
	300	**<0.0001**	**<0.0001**	**0.0007**	0.2460	—
	500	**<0.0001**	**<0.0001**	**0.0007**	0.2460	1.0000
ML	30	—	—	—	—	—
	50	0.9761	—	—	—	—
	100	1.0000	0.9761	—	—	—
	200	0.0531	**0.0075**	0.0531	—	—
	300	**<0.0001**	**<0.0001**	**<0.0001**	0.0531	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	**0.0075**	0.9761
VT	30	—	—	—	—	—
	50	0.9754	—	—	—	—
	100	0.6756	0.9754	—	—	—
	200	**<0.0001**	**0.0007**	**0.0070**	—	—
	300	**<0.0001**	**<0.0001**	**0.0007**	0.9754	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	0.6756	0.9754

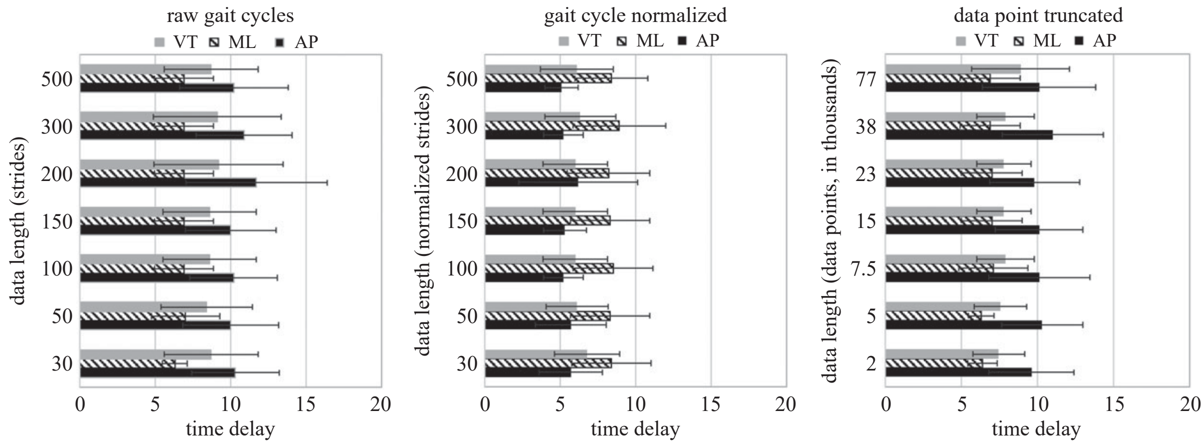

Figure 2. Mean (s.d.) of time delay when calculated with different data lengths and preprocessing methods for every signal direction: vertical (VT), mediolateral (ML) and anteroposterior (AP).

Table 6. Post hoc Tukey test results comparing embedding dimensions at all data lengths for the gait data processed using method 3 (fixed number of data points).

		data length comparison for method 3				
	data length (gc)	30	50	100	200	300
AP	30	—	—	—	—	—
	50	0.9684	—	—	—	—
	100	0.3766	0.8438	—	—	—
	200	**<0.0001**	**<0.0001**	**<0.0001**	—	—
	300	**<0.0001**	**<0.0001**	**<0.0001**	**0.0114**	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	**0.0003**	0.9373
ML	30	—	—	—	—	—
	50	0.9684	—	—	—	—
	100	0.3766	0.8438	—	—	—
	200	**<0.0001**	**<0.0001**	**<0.0001**	—	—
	300	**<0.0001**	**<0.0001**	**<0.0001**	**0.0114**	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	**0.0003**	0.9373
VT	30	—	—	—	—	—
	50	0.9793	—	—	—	—
	100	0.7104	0.9793	—	—	—
	200	0.0668	0.2854	0.7104	—	—
	300	0.0668	0.2854	0.7104	1.0000	—
	500	**<0.0001**	**<0.0001**	**<0.0001**	0.0668	0.0668

The majority of publications that calculate the LyE for gait use a single time delay for every subject [35]. Although one paper has looked at some of the differences between individualized and a pre-selected fixed time delay, the researchers were specifically investigating the intra-patient reliability of LyE [12] and only in the ML direction. A more in-depth study into how underestimating or overestimating the τ in the LyE calculation is needed to understand its importance and contribution to the reliability of the LyE for gait.

We discovered that the calculated d_E varies with respect to data length. However, a steady state d_E can be reached as long as the minimum data requirement is met for the dynamical system. If we look at the simulated systems, shown in the electronic supplementary material, figure S2, the calculated Lorenz quickly reaches a consistent d_E after 5×10^3 in the x-direction and 2×10^3 points in the y- and z-direction. The Rössler system required 5×10^3 points to reach a steady state d_E in the x time series, 2×10^3 points in the y, and 500 points in the z. The z time series did not always converge on to the same d_E as the other time series, however. This could be a sign that the z time series has insufficient information in its signal to be used for phase space reconstruction.

The gait data also reached a steady-state embedding dimension; however, it required at least 300 gait cycles (figure 3). No significant differences were found between 300 and 500 gait cycles for all signal directions (table 2);

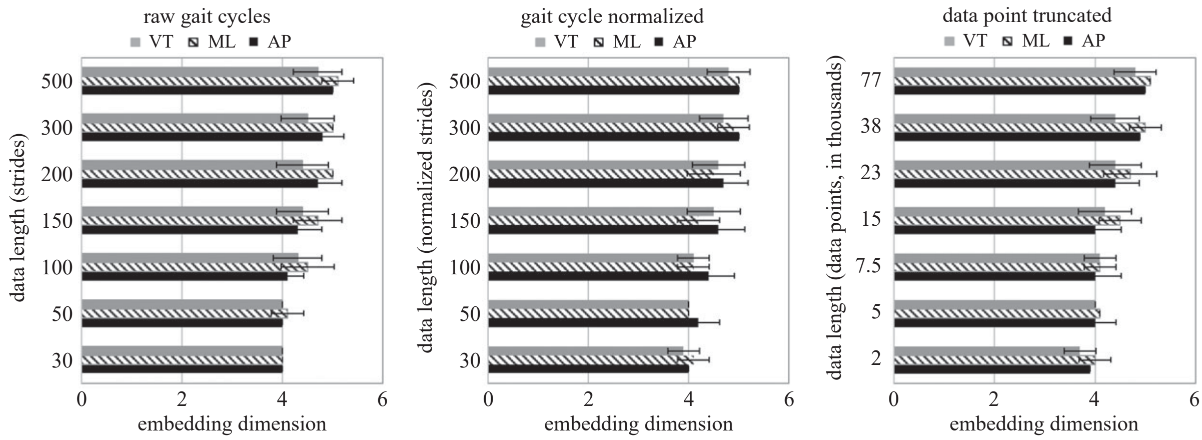

Figure 3. Mean (s.d.) of embedding dimension when calculated with different data lengths and preprocessing methods for every signal direction: vertical (VT), mediolateral (ML) and anteroposterior (AP).

however, this relationship changes when you independently test changes in d_E with respect to data length for each preprocessing method. Preprocessing method 1 (fixed number of gait cycles) required the least amount of data to reach a steady state d_E. Steady state was reached at 200 gait cycles in the AP and ML directions and 100 gait cycles in the VT direction. For preprocessing method 2 (fixed number of stride normalized gait cycles), a steady state d_E is reached at 300 gait cycles in the ML direction and 200 gait cycles in the AP and VT directions. Preprocessing method 3 (fixed number of data points) had the highest data requirements for finding a stable d_E, requiring 300 gait cycles in the AP and ML directions, and 200 gait cycles in the VT direction. It is important to note that all preprocessing methods, in every acceleration direction, did converge onto a d_E of 5 after 300 gait cycles. Therefore, we find that an embedding dimension of 5 is sufficient for processing young healthy adult gait data for LyE. Future research should look at how much d_E affects the final outcome of LyE using either algorithm for calculating local dynamic stability.

This study had some limitations with using only young healthy adults. We cannot assume that data length will have similar effects on τ and d_E when looking at different population groups, e.g. healthy or frail elderly. Although a treadmill was used for this experiment, this should not have an impact on the reconstruction of the phase space itself. The calculated LyE is believed to be different from treadmill and over-ground walking owing to slightly different gait dynamics used to adapt to each situation [28,36]. However, this terrain difference has no influence on the method of phase space reconstruction.

The current study provided novel information by systematically investigating the effect of data length on time delay and embedding dimension in gait data. Data length does not play a large role in the calculation of τ using AMI, while a minimum data requirement must be first met when calculating the d_E using FNN. Therefore, the differences in the LyE at various data lengths are not owing to the reconstruction of the gait attractor, but more likely owing to the increasing signal-to-noise ratio as the data length increases when calculating the LyE. Additionally, we investigated the effect of three methods of gait data preprocessing. We found method choice significantly impacts the value of the τ but not the d_E when sufficient data are provided. We recommend using at least 200 gait cycles for calculating the d_E when using a preprocessing method similar to method 1 and 300 gait cycles when using methods 2 or 3. However, this might not be necessary as this analysis revealed that young healthy gait data can be processed using a d_E of 5 for any acceleration data regardless of how the data are preprocessed.

Ethics. This study was approved by the Institutional Review Board of Arizona State University (ASU FWA 00009102).

Data accessibility. The electronic supplementary material including data and MATLAB scripts for this manuscript is available at: https://github.com/ASU-LocomotionResearchLab/ParamReconstructionEffects.

Authors' contributions. V.S.H. and T.E.L. conceived and planned the experiment. V.S.H. acquired and analysed the data. V.S.H., M.L.S. and T.E.L. contributed to the interpretation of the results. V.S.H. took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Competing interests. We have no competing interests.

Funding. This research was supported by ARCS Foundation scholarship to Victoria A. Smith Hussain. Thurmon Lockhart is supported by BRAIN Project: Custom Designed Wearable Sensor System for Fall Risk Assessment and by Michael J. Fox Foundation: Protective Step Training in People with PD and Postural Disturbance. This research was supported by NSF-Information and Intelligent Systems (IIS) and Smart and Connected Health (1065442, 1065262 and 1547466).

References

1. Grebogi C, McDonald SW, Ott E, Yorke JA. 1983 Final state sensitivity: an obstruction to predictability. *Phys. Lett. A* **99**, 415–418. (doi:10.1016/0375-9601(83)90945-3)
2. Baker GL, Gollub JP. 1996 *Chaotic dynamics, an introduction*, 256 p, 2nd edn. New York, NY: Cambridge University Press.
3. Grebogi C, Ott E, Yorke JA. 1987 Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics. *Science* **238**, 632–638. (doi:10.1126/science.238.4827.632)
4. Lorenz EN. 1963 Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141. (doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
5. Thompson JMT, Stewart HB. 1986 *Nonlinear dynamics and chaos*. Chichester, UK: John Wiley & Sons, Ltd.
6. Toebes MJP, Hoozemans MJM, Furrer R, Dekker J, Van Dieën JH. 2012 Local dynamic stability and variability of gait are associated with fall history in elderly subjects. *Gait Posture* **36**, 527–531. (doi:10.1016/j.gaitpost.2012.05.016)

7. Lockhart TE, Liu J. 2008 Differentiating fall-prone and healthy adults using local dynamic stability. *Ergonomics* **51**, 1860–1872. (doi:10.1080/00140130802567079)
8. Fino PC, Mancini M, Curtze C, Nutt JG, Horak FB. 2018 Gait stability has phase-dependent dual-task costs in Parkinson's disease. *Front Neurol.* **9**, 1–10. (doi:10.3389/fneur.2018.00001)
9. Speedtsberg MB, Christensen SB, Stenum J, Kallemose T, Bencke J, Curtis DJ, Jensen BR. 2018 Local dynamic stability during treadmill walking can detect children with developmental coordination disorder. *Gait Posture* **59**, 99–103. (doi:10.1016/j.gaitpost.2017.09.035)
10. Bruijn SM, van Dieën JH, Meijer OG, Beek PJ. 2009 Is slow walking more stable? *J. Biomech.* **42**, 1506–1512. (doi:10.1016/j.jbiomech.2009.03.047)
11. England SA, Granata KP. 2007 The influence of gait speed on local dynamic stability of walking. *Gait Posture* **25**, 172–178. (doi:10.1016/j.gaitpost.2006.03.003)
12. van Schooten KS, Rispens SM, Pijnappels M, Daffertshofer A, Van Dieën JH. 2013 Assessing gait stability: the influence of state space reconstruction on inter- and intra-day reliability of local dynamic stability during over-ground walking. *J. Biomech.* **46**, 137–141. (doi:10.1016/j.jbiomech.2012.10.032)
13. Terrier P, Reynard F. 2014 To what extent does not wearing shoes affect the local dynamic stability of walking?: effect size and intrasession repeatability. *J. Appl. Biomech.* **30**, 305–309. (doi:10.1123/jab.2013-0142)
14. Bruijn SM, van Dieën JH, Meijer OG, Beek PJ. 2009 Statistical precision and sensitivity of measures of dynamic gait stability. *J. Neurosci. Methods* **178**, 327–333. (doi:10.1016/j.jneumeth.2008.12.015)
15. Kang HG, Dingwell JB. 2009 Dynamic stability of superior vs. inferior segments during walking in young and older adults. *Gait Posture* **30**, 260–263. (doi:10.1016/j.gaitpost.2009.05.003)
16. Cignetti F, Decker LM, Stergiou N. 2012 Sensitivity of the Wolf's and Rosenstein's algorithms to evaluate local dynamic stability from small gait data sets: response to commentaries by Bruijn *et al*. *Ann. Biomed. Eng.* **40**, 2507–2509. (doi:10.1007/s10439-012-0665-6)
17. Broomhead DS, King GP. 1986 Extracting qualitative dynamics from experimental data. *Phys. D Nonlinear Phenom.* **20**, 217–236. (doi:10.1016/0167-2789(86)90031-X)
18. Fraser AM, Swinney HL. 1986 Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**, 1134–1140. (doi:10.1103/PhysRevA.33.1134)
19. Kennel MB, Brown R, Abarbanel HDI. 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403–3411. (doi:10.1103/PhysRevA.45.3403)
20. Stenum J, Bruijn SM, Jensen BR. 2014 The effect of walking speed on local dynamic stability is sensitive to calculation methods. *J. Biomech.* **47**, 3776–3779. (doi:10.1016/j.jbiomech.2014.09.020)
21. Myers SA, Pipinos II, Johanning JM, Stergiou N. 2011 Gait variability of patients with intermittent claudication is similar before and after the onset of claudication pain. *Clin. Biomech.* **26**, 729–734. (doi:10.1016/j.clinbiomech.2011.03.005)
22. Van Schooten KS, Rispens SM, Elders PJM, van Dieën JH, Pijnappels M. 2014 Toward ambulatory balance assessment: estimating variability and stability from short bouts of gait. *Gait Posture* **39**, 695–699. (doi:10.1016/j.gaitpost.2013.09.020)
23. Dingwell JB, Marin LC. 2006 Kinematic variability and local dynamic stability of upper body motions when walking at different speeds. *J. Biomech.* **39**, 444–452. (doi:10.1016/j.jbiomech.2004.12.014)
24. Raffalt PC, Kent JA, Wurdeman SR, Stergiou N. 2019 Selection procedures for the largest Lyapunov exponent in gait biomechanics. *Ann. Biomed. Eng.* **47**, 913–923. (doi:10.1007/s10439-019-02216-1)
25. Treutwein B. 1995 Adaptive psychophysical procedures. *Vision Res.* **35**, 2503–2522. (doi:10.1016/0042-6989(95)00016-X)
26. Pan J, Tompkins WJ. 1985 Real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **BME-32**, 230–236. (doi:10.1109/TBME.1985.325532)
27. Norris M, Kenny IC, Anderson R. 2016 Comparison of accelerometry stride time calculation methods. *J. Biomech.* **49**, 3031–3034. (doi:10.1016/j.jbiomech.2016.05.029)
28. Dingwell JB, Cusumano JP, Cavanagh PR, Sternad D. 2001 Local dynamic stability versus kinematic variability of continuous overground and treadmill walking. *J. Biomech. Eng.* **123**, 27. (doi:10.1115/1.1336798)
29. Rispens SM, Pijnappels M, van Dieën JH, Van Schooten KS, Beek PJ, Daffertshofer A. 2014 A benchmark test of accuracy and precision in estimating dynamical systems characteristics from a time series. *J. Biomech.* **47**, 470–475. (doi:10.1016/j.jbiomech.2013.10.037)
30. Rosenstein MT, Collins JJ, De Luca CJ. 1993 A practical method for calculating largest Lyapunov exponents from small data sets. *Phys. D Nonlinear Phenom.* **65**, 117–134. (doi:10.1016/0167-2789(93)90009-P)
31. Wolf A, Swift JB, Swinney HL, Vastano JA. 1985 Determining Lyapunov exponents from a time series. *Phys. D Nonlinear Phenom.* **16**, 285–317. (doi:10.1016/0167-2789(85)90011-9)
32. Rossler OE. 1976 An equation for continuous chaos. *Phys. Lett. A* **57**, 397–398. (doi:10.1016/0375-9601(76)90101-8)
33. Abarbanel HDI, Brown R, Sidorowich JJ, Tsimring LS. 1993 The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.* **65**, 1331–1392. (doi:10.1103/RevModPhys.65.1331)
34. Kantz H, Schreiber T. 2004 *Nonlinear time series analysis*, 2nd edn. Cambridge, UK: Cambridge University Press.
35. Mehdizadeh S. 2018 The largest Lyapunov exponent of gait in young and elderly individuals: a systematic review. *Gait Posture* **60**, 241–250. (doi:10.1016/j.gaitpost.2017.12.016)
36. Terrier P, Dériaz O. 2011 Kinematic variability, fractal dynamics and local dynamic stability of treadmill walking. *J. Neuroeng. Rehabil.* **8**, 12. (doi:10.1186/1743-0003-8-12)