Author for correspondence:
Alessandro Maria Selvitella
e-mail: aselvite@pfw.edu

THE ROYAL SOCIETY PUBLISHING

On the variability and dependence of human leg stiffness across strides during running and some consequences for the analysis of locomotion data

Alessandro Maria Selvitella[1,2,3] and Kathleen Lois Foster[3,4]

[1]Department of Mathematical Sciences, Purdue University Fort Wayne, 2101 East Coliseum Boulevard, Fort Wayne, IN 46805, USA
[2]eScience Institute, University of Washington, 3910 15th Avenue Northeast, Seattle, WA 98195, USA
[3]NSF-Simons Center for Quantitative Biology, Northwestern University, 2200 Campus Drive Evanston, IL 60208, USA
[4]Department of Biology, Ball State University, 2000 West University Avenue, Muncie, IN 47306, USA

AMS, 0000-0002-1620-998X; KLF, 0000-0002-7365-7885

Typically, animal locomotion studies involve consecutive strides, which are frequently assumed to be independent with parameters that do not vary across strides. This assumption is often not tested. However, failing in particular to account for dependence across strides may cause an incorrect estimate of the uncertainty of the measurements and thereby lead to either missing (overestimating variance) or over-evaluating (underestimating variance) biological signals. In turn, this impacts replicability of the results because variability is accounted for differently across experiments. In this paper, we analyse the changes of a couple of measures of human leg stiffness across strides during running experiments, using a publicly available dataset. A major finding of this analysis is that the time series of these measurements of stiffness show autocorrelation even at large lags and so there is dependence between individual strides, even when separated by many intervening strides. Our results question the practice in biomechanics research of using each stride as an independent observation or of sub-selecting strides at small lags. Following the outcome of our analysis, we strongly recommend caution in doing so without first confirming the independence of the measurements across strides and without confirming that sub-selection does not produce spurious results.

1. Introduction

In biomechanical studies of human or animal locomotion, in particular in the analysis of terrestrial gaits, it is common practice for researchers to take measurements and compute variables of interest for each stride and then average or analyse the values of those variables across strides, often considering each stride an independent and identically distributed (iid) observation [1–5]. As a general rule, an effort is made to keep all possible strides available. Most of the experiments in the comparative and human locomotion literature are based on sample sizes of less than 50, and often even smaller, less than 10 (e.g. [6–10]), and so researchers want to keep as much information as possible from the experiments performed. Enlarging the study is often either costly or not feasible because of characteristics of the species (e.g. rapid growth rate, challenging/rare behaviours) or the conditions of the equipment or because the population to be sampled does not allow for larger testing procedures. For example, enlarging a sample for a comparative study could mean going back to the field and taking a new sample, a very cumbersome path to follow [11]. Extending the data collection for longer periods of time would not be a solution as that would require complicating the model and, therefore, the extra inflow of data may not keep up with the extra-complexity required to model a growth study. These small sample issues for biomechanical studies are an ubiquitous feature of the field, even in this big data era. Note that the practice of treating the strides as independent is diffuse across different areas of the locomotion literature. Researchers studying humans [3,6,12–14], bipedal birds [5,8,10,15,16], lizards [9,17–20] and other animals as well [2,4,21–23], estimate parameters of biomechanical relevance by averaging measurements across consecutive locomotor cycles and using statistical methods based on the independence hypothesis.

Sometimes, to avoid the extra challenges posed by correlated data, researchers choose to only use measurements from every two or three strides for their average estimate [12,24]. In extreme cases, only one single stride per trial is considered [25–27]. This strategy has distinct disadvantages: it reduces the total number of strides that can be used in the study and disrupts the natural correlation structure present in the data, between strides. In many cases, researchers do not rely on the structure of the data to decide how many strides to drop or how many can be considered independent and, for those which are not independent, to determine which type of dependence structure there is [12,24–26]. When steady-state locomotion is being studied, the first and last strides from a given trial are often excluded or take place off-camera and so are not recorded [1]. Such decisions to exclude strides are, in general, made by the single researcher and they are not standardized across laboratories. Using personalized stride exclusion criteria does not only risk to preclude the correct reporting of the results of the experiments, to miss the opportunity to learn about the irregular biomechanics at the beginning of a cyclical motion, and to jeopardize the replicability of the experiments across laboratories, but also is questionable from the data science perspective. It is hard to justify the decision to exclude some strides from analysis simply because they might be dependent, particularly considering that any dependence present in a time series may arguably be the most important feature.

The main objectives of this manuscript are, first, to underline that it is incorrect to not account for variability of measurements across strides in the analysis of data from locomotion experiments and, then, discuss the possible consequences of doing so. For these objectives, we concentrate on one relevant example and study it extensively, but we believe that our considerations are valid in general for terrestrial locomotion experiments which involve multi-stride data. We study the variability of human leg stiffness across strides during running on a treadmill at different velocities and use a publicly available dataset [28] to illustrate our argument. Note that this topic has not received extensive attention in human locomotion studies and so our results also provide new insights on this important problem.

Most commonly, human running is modelled as a *spring-mass/spring loaded inverted pendulum* (SLIP) [29–32]. This model is reductionist and tries to extract the most fundamental features of the complex dynamics of human running with a low-dimensional system of ordinary differential equations describing bouncing gaits. The resulting model has qualitative validity in the sense that trajectories of the SLIP system resemble those of bipedal runners and many other multi-pedal runners ([29,30,33] and the large body of literature that cites these papers). As far as we know, there are no proven mathematical theorems that guarantee under which regimes (e.g. small joint angles, low horizontal velocity) the high-dimensional dynamical system describing an animal running can be accurately determined by a low-dimensional model, such as the two-dimensional SLIP system. One of the consequences of the lack of guarantees of these reductionist models is the difficulty in establishing the best methodology for estimating parameters of interest, such as leg stiffness, through models like

the SLIP model [34–38]. The basic idea of this *spring-mass* model is to use a linear spring that follows Hooke's law and compute the stiffness by dividing the maximum (max) vertical ground reaction force (GRF), typically measured using kinetic data from force platforms, and the compression of the leg deduced using kinematic data, such as the position of the centre of mass (CoM) [29,30].

Considerable discussion in the literature has centred around how to calculate leg stiffness during human running and the validity of the various methods has been questioned [34]. Some methods include the measurements of the leg length and GRFs at touch down, such as those employed in [35,36,39], and the use of a mechanical arm to track the movement of the CoM [40]. In some situations, the stiffness is a parameter estimated from a chosen model and, in such cases, it simply comes from the model assumptions rather than existing as a biological parameter itself. In the SLIP model, the stiffness represents some sort of summary measure of the properties of the individual which appears to globally possess a *spring-mass* behaviour, but it is not a direct measurement of the elasticity of any specific muscle/tendon. In [41], the authors noted that the stiffness of the leg cannot be calculated directly neither in humans nor in bipedal birds and that an effective leg stiffness can be estimated only if assuming that the entire body behaves like a spring-mass system during running. It does make sense to estimate such a parameter from the data, but the interpretation of such an estimate needs to be used with caution. As mentioned, there are many methods used to estimate stiffness, some of them described in [42] (see also [43]), in which authors compare five methods for the estimation of the stiffness parameter based on kinematic and kinetic data. The authors in [34] noticed that reductionist models often require untested assumptions for their mathematical analysis to be carried out, but that those assumptions are not tested and comparisons with the consequences derived from different assumptions are rarely made. Furthermore, [44] noted that the method in [30] underestimated leg stiffness compared to their own method and that, in [30], the authors did not note a dependence of leg stiffness with velocity, whereas [42] reported similar leg stiffness values to those following the method proposed in [30]. Also, the authors in [34] compared leg stiffness values measured using direct kinematic and kinetic methods with leg stiffness values calculated using the most common methods present in the literature. They found that stiffness estimates are highly variable across methods and studies. In addition to the challenge of estimating parameters of interest, a further difficulty comes from the fact that, in most models, the human body is assumed to be rigid, but GRFs do not directly act on the CoM of the body. Such a simplification neglects dissipative forces in the transmission of the GRFs from the centre of pressure of the foot in contact with the ground to the CoM. There is actually not a strong argument to assume that the GRFs act on the CoM directly [45].

The authors in [46] found that human runners adjust their leg stiffness [36,47] to accommodate changes in surface stiffness, allowing them to run with similar dynamics on different surfaces. Several studies have investigated the way in which leg stiffness is adjusted to accommodate changes in ground level. In particular, the authors in [48] studied how runners modulate ankle and knee joint stiffness, finding, among their other results, that the ankle joint stiffness depends on the vertical height of a step, similar to the global leg stiffness. In [49], the author addressed the problem of joint level compliance during human walking and how joint stiffness is modulated during human walking on flat surfaces, inclined surfaces, and stairs. The results found that stiffness estimation was much lower than those found in running ([50]; see also [51,52]. As far as we know, there has not been an extensive study of how stiffness changes across strides at different running velocities. Leg stiffness is also of interest in animal biomechanics [34,53–56] (and papers referring to these) and the variability of leg stiffness as a function of speed has been a source of discussion in comparative biomechanics as well. Some experimental studies have reported that animal leg stiffness is independent of speed [36,46,57].

It is a very hard problem to determine how age influences the changes across strides of leg stiffness while running at different velocities [58]. Even when considering a single individual, there may be many confounding factors (e.g. history of injury, typical footwear etc.) which need to be taken into consideration to properly address the problem. Note that being able to understand how the dynamics of older versus younger individuals changes, in particular how leg stiffness is modulated through stride as a function of age, can have important consequences in fields such as kinesiology, rehabilitation and prosthetics. With longer life expectancy, many more people encounter orthopaedic problems [59], including those with moving, particularly those related to optimal running modes. How these problems arise is still poorly understood. Muscle stiffness has been reported to be associated with the reduction of force generation capability of quadriceps in older populations [60]. Such a decrease in muscle strength has been associated with the slowing of movements in elderly people, together with this rigidity or increased stiffness of muscles ([60]; see also the work in [58,61]. None of these results seem to address if there is a significant change of leg stiffness through strides or prolonged running experiments across age groups. The problem

of how changes in stiffness across strides depend on mass is complex as well, as there are no simple physical principles that help in the modelling of such a change. Many experiments and analyses need to be done. Some studies have investigated how leg stiffness scales with respect to mass in bipedal and quadripedal mammals in both running and hopping, using experimental and simulated data [35,62–65], but, as far as we know, none have discussed how mass impacts the change in stiffness across strides.

In summary, the goals of the paper are the following: (i) to determine the influence that neglecting the dependence structure of strides can have on the results from the analysis of kinetic and kinematic data, using measurements of leg stiffness averaged across strides as a running example; and (ii) to provide a better understanding of the variability across strides of human leg stiffness during treadmill running.

2. Methods

In this section, we describe the methods we used for our analysis.

2.1. Dataset

The dataset that we used is publicly available [28]. As explained in the data repository [28], the dataset collects multivariate time series of the trajectories of signals from sensors applied to the human body while running on a treadmill at $2.5 \, \mathrm{m \, s^{-1}}$, $3.5 \, \mathrm{m \, s^{-1}}$ and $4.5 \, \mathrm{m \, s^{-1}}$. The time series include both kinematics and kinetics of 28 subjects and also includes metadata with age and mass among the variables considered. Although, we will analyse only the dynamics in the sagittal plane, the data comprise three-dimensional coordinates collected from a motion-capture system.

2.2. Pre-processing of the data

Before starting the analysis, we pre-processed the data. The frequency of collection of kinematic and kinetic measurements were different; the time series of the forces were collected at 300 Hz while the time series of the markers were collected at 150 Hz. For our analysis, we sub-selected every other datapoint in the ground reaction force data and paired them with the full kinematic data. During the aerial phase, the vertical components of the forces exerted on the foot is zero, and so the only force applied to the CoM is the gravitational force. This is true because the foot is not touching the treadmill and so force platforms cannot record any force. We excluded the aerial phase from our analysis, as we are interested in understanding the elastic properties of the leg and in the aerial phase those elastic forces are zero. To do so, we extracted from the full kinematic and kinetic time series of each subject the measurements of the forces for all instants constituting only the stance phases as determined by the condition that the vertical force recorded was positive ($F_y > 0$). Furthermore, the kinematic data consisted of several time series of markers placed on the body of the participants recorded during the experiments. Not all of these markers were useful for understanding the reaction of the individual (e.g. change in leg stiffness) to the collision of the foot onto the treadmill. As we only needed the information about the CoM, which is located approximately at the height of the hips, we retained and averaged only the coordinates of the left and right markers (in the sagittal plane with positive x concordant with the running direction and positive y in the direction foot-to-head) that were closest to this position: the *anterior superior Iliac spine*, the *posterior Iliac spine* and the *Iliac crest* coordinates.

Then, we extracted each single stride and stored the time series of these strides in separate variables. Finally, the coordinate system in which the dataset was collected was centred at an origin static with respect to the treadmill. Therefore, we performed a change of variables that allowed us to recentre the data: $x' = x + v^* t$, $y' = y + v^* t$ with (x, y) the coordinates with respect to the treadmill origin, t the time index and v the velocity of the treadmill in that trial. Note that the way in which we computed the CoM is just one of the ways available in the literature. For example, the authors in [66] estimated the vertical and relative adaptation of the CoM using three different markers: the coordinates of the fifth lumbar vertebrae and the seventh cervical vertebrae and the average of the coordinates of the right and left greater trochanter from the upper leg.

2.3. Calculations of leg stiffness

Leg stiffness can generally be defined as $k = \partial F / \partial L$, with F being the vertical component of the GRF and L the leg length during the stance phase [42]. As mentioned in [42,43], there are many ways to define that quantity,

both in a linear and nonlinear way. Here, we do not argue in favour or against any of the methods, but we consider a sub-selection of those methods. We compute the leg stiffness in two different ways:

— *method 1*: the maximal leg stiffness k_{max} was computed as the resultant force in the direction of the leg spring F_{leg} divided by the leg compression $\Delta L : k_{leg} = F_{leg}/\Delta L$ (see for example, [34]). F_{leg} was computed as the value of the vertical GRF, F_y, computed at the lowest height y of the centre of mass, after correcting F_y for the weight of the individual. The leg compression ΔL was computed as:

$$\Delta L = \max(y) - \min(y) + l_0 * \left(1 - \sqrt{1 - \left(\frac{\max(x) - \min(x)}{(2*l_0)}\right)^2}\right).$$

Here, max (y), min (y) are the maximal and minimal vertical position, with respect to the height of the centre of mass during the stance phase, while max (x), min (x) are the maximal and minimal x coordinates of the centre of mass during that particular stance phase and l_0 is the height of the CoM at rest. Here, x and y are calculated after the change of variable performed in the pre-processing step; and

— *method 2*: we computed k_{OLS} using the ordinary least-squares method for a linear model without intercept. The independent variable was the GRF defined using a *spring-mass* model:

$$\text{GRF}_y = y * \left(\frac{l_0}{\sqrt{x^2 + y^2} - 1}\right),$$

and the dependent variable was the vertical component of the recorded GRF plus the weight ($m*g$, with $g = 9.81$ m s^{-2}) of the individual.

2.4. Hypothesis tests for stride independence

In this subsection, we describe the procedures that we followed to test the hypothesis H_0: *strides are independent* versus H_a: *strides are not independent*, and in particular the hypothesis that the leg stiffness measurements k_{OLS}, k_{max} are or not stride independent. Note that if summary statistics calculated from different populations are dependent, then the two populations are dependent as well [67]. Therefore, the dependence of the stiffness on stride indicates a dependence of the stance phases themselves within experimental trials. We concentrated on tests which measure the possibility of autocorrelation in the time series of the stiffness measurements. All tests were performed at significance level $\alpha = 0.05$.

Yule test

The simplest method to test for autocorrelation in a time-series dates back to Yule [68,69]. For a given sample $Y_1, \ldots, Y_n$, the population autocorrelation function (ACF) $\rho_Y(k)$ is estimated by the sample ACF $\hat{\rho}_Y(k)$ for $k = 1, 2, \ldots$. Note that the ACF $\hat{\rho}_Y(k)$ is well defined for all t, $k = 1, 2, \ldots$, even for non-stationary processes (but with finite first and second moments). The ACF is given by

$$\hat{\rho}_Y(k) := \frac{\hat{\gamma}_Y(k)}{\hat{\gamma}_Y(0)}, \quad \hat{\gamma}_Y(k) = \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \bar{X}_T)(X_{t+k} - \bar{X}_n), \quad t = 0, 1, \ldots, n-1.$$

The asymptotic distribution of the sample autocorrelations $\hat{\rho}_Y(k)$ for various lags $k = 1, 2, \ldots$ for a general stationary linear process [70,71] takes the form

$$\sqrt{n}(\hat{\rho}_Y(1) - \rho_Y(1), \ldots, \hat{\rho}_Y(m) - \rho_Y(1)) \to \mathcal{N}(0, W),$$

as $n \to +\infty$ with W given by Bartlett's formula [72], which, in the case of a null hypothesis given by iid random variables, reduces to the identity matrix $W = Id_m$ and so

$$\sqrt{n}(\hat{\rho}_Y(1), \ldots, \hat{\rho}_Y(m)) \to \mathcal{N}(0, Id_m),$$

with Id_m the identity matrix in dimension m and m, the maximum lag considered. Therefore, to test the hypothesis

$$H_0 : \rho_k = 0 \ vs \ H_A : \text{ otherwise},$$

at level α, it is enough to verify that $|\hat{\rho}_Y(k)| > z_{\alpha/2}/\sqrt{n}$ with n the size of the sample observed [73].

There is a vast literature on testing for significant autocorrelation in time series, starting from tests not only at one single lag k, but for a series of lags, such as $H_0 : \rho_Y(1) = \ldots = \rho_Y(m) = 0$ [74,75]. We considered these cases too.

Ljung-Box test for independence

The Yule test detects if there are correlations at single lags. Sometimes single autocorrelations oscillate and dependency in a time series can be detected only by grouping autocorrelations together and verifying their magnitudes. The Ljung-Box test [75] aims at detecting if a group of autocorrelations of a time series is significantly different from zero. The Ljung-Box test statistic is given by:

$$Q = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_Y(k)^2}{n-k},$$

where n is the sample size and m is the number of lags being tested. We performed a Ljung-Box test with maximum lag 40 at each velocity and for both measures of stiffness, k_{max} and k_{OLS}. We collected the p-values for each of the tests. We considered a Bonferroni-type adjustment with significance level α/N with $\alpha = 0.05$ and $N = 28$. As a follow-up to the Ljung-Box test, we used linear models to determine how the dependency of stride was influenced by age and mass, using the p-values of the Ljung-Box test for maximum lag equal to 40 as the dependent variable and age and mass as independent variables.

Linear mixed effects models

We analysed how the stiffness was influenced by stride and velocity (see appendix A 4). To do so, we developed several linear mixed effects models (LMEMs) with subject dependency as a random effect [76]. We developed LMEMs including the velocity as a fixed effect, but also separate models for each single velocity. We restricted our attention to the case where $\beta_{1,j} = \beta_1$, namely the case in which only the intercept of the model includes random effects accounting for between subject variations. More specifically, we considered the following models for both $k = k_{max}$ and $k = k_{OLS}$:

$$k_{ij} = \beta_{i,0} + \beta_1 S_{ij} + \varepsilon_{ij},$$

for each of the three velocities ($v = 2.5$ m s^{-1}, $v = 3.5$ m s^{-1} and $v = 4.5$ m s^{-1}). Here, $j = 1, \ldots, 28$ and $i = 1, \ldots, n_j$ with n_j the number of strides for subject j for each single velocity. S_{ij} represents stride number j for subject i. Furthermore, we considered the models with velocity included for both $k = k_{max}$ and $k = k_{OLS}$:

$$k_{ij} = \beta_{i,0} + \beta_1 S_{ij} + \beta_2 v_{ij} + \varepsilon_{ij},$$

for each of the three velocities. As before, $j = 1, \ldots, 28$ and $i = 1, \ldots, n_j$ with n_j the number of strides for subject j across velocities. Larger models including random effects affecting the inclination were not considered because, in such models, the estimation procedure was unstable, given that the complexity of those models was too high with respect to the sample size available in this study. Parameter estimation was performed using the restricted maximum likelihood criterion [77,78]. We fitted all LMEMs using the R-package *lmer* (see [79]).

3. Results

In this section, we outline the results of our analysis.

3.1. Dependence of stiffness across strides within individual subjects

In all of the six conditions analysed (two stiffness measures and three velocities), the Yule test revealed the presence of significant autocorrelation in the stiffness measurements that persists across strides even for lags larger than 5 in 5/28, 4/28, 6/28, 7/28, 8/28 and 8/28 subjects (figure 1, panels *a–f*, respectively). With the notation 'X/Y', we mean 'X out of Y'. Note that a large number of locomotion studies consider each stride as independent and so, even just correlation at lag 1 would be a significant finding. The Ljung-Box test also found significant dependence across strides for many individuals. It determined that 11/28, 12/28, 8/28, 13/28, 16/28 and 6/28 individuals (figure 2, panels *a–f*, respectively) showed a dependence in stiffness across strides (only for those up to 40 lags apart) and so the strides are not iid; in particular they are not independent.

3.2. Dependence of strides with age and mass

The analysis of age dependence of the Ljung-Box test indicates an increased dependence across strides with increasing age, but the estimates show high variability and none of the tests rejected the null
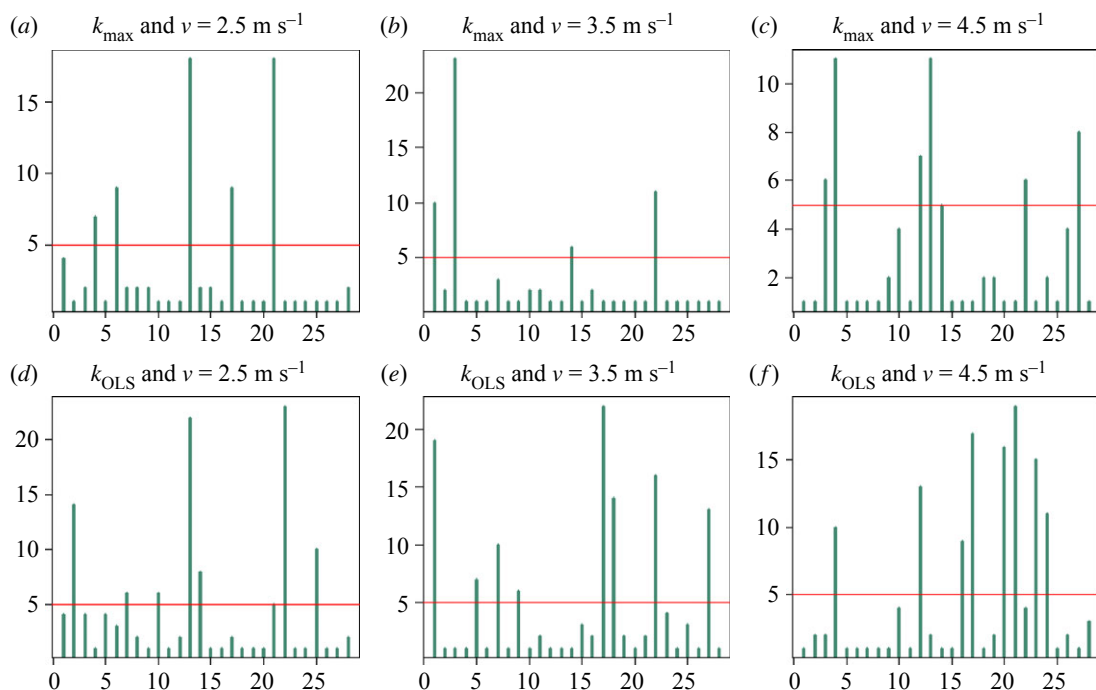
Figure 1. Yule test showing autocorrelation across strides at $v = 2.5$ m s^{-1} (a,d), $v = 3.5$ m s^{-1} (b,e), and $v = 4.5$ m s^{-1} (c,f) for k_{max} (a–c) and k_{OLS} (d–f). The y-axis of each plot corresponds to the lags, while the x-axis indexes the subjects. The vertical bars represent the largest lag which shows correlation across strides for each of the 28 subjects, while the horizontal line is fixed at a lag of 5 for illustrative purposes. In each of the conditions and for both k_{max} and k_{OLS}, there is correlation across strides for a relevant number of subjects.
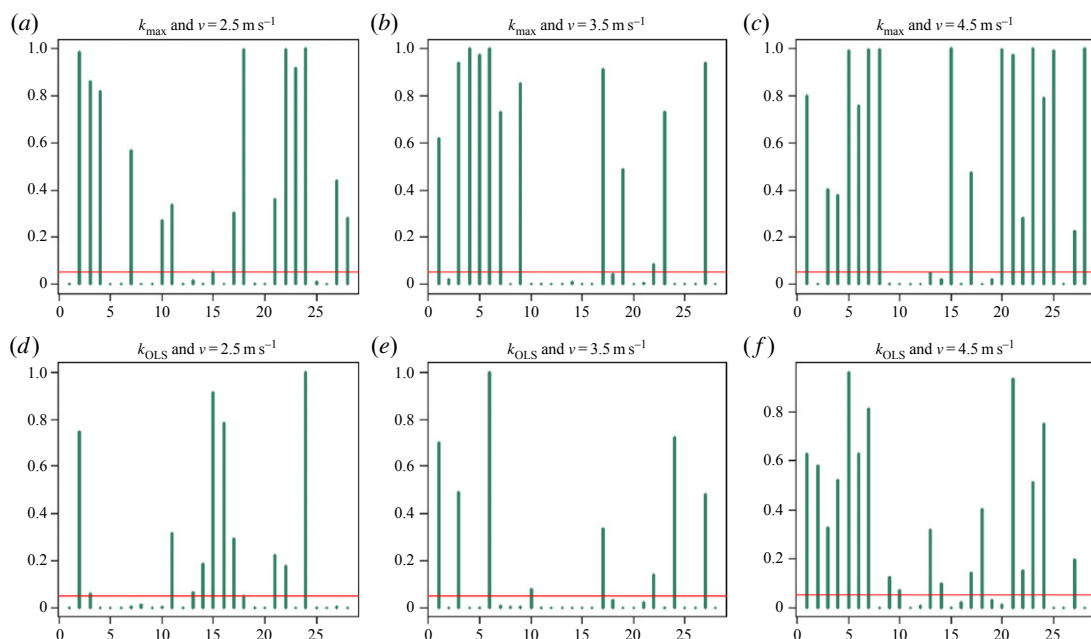

Figure 2. Ljung-Box test showing dependence across strides for the three velocities, $v = 2.5$ m s^{-1} (a,d), $v = 3.5$ m s^{-1} (b,e), and $v = 4.5$ m s^{-1} (c,f), and the two measures of stiffness, k_{max} (a–c) and k_{OLS} (d–f). The y-axis of each plot corresponds to the Ljung-Box p-value, while the x-axis indexes the subjects. The vertical bars represent the p-value of the Ljung-Box test for maximum lag equal to 40 for that subject, while the horizontal line is fixed at the level of significance ($\alpha = 0.05$) for illustrative purposes.

hypothesis of p-value independent of age ($p = 0.3574, 0.7737, 0.3982, 0.6224, 0.3203$ and 0.8704 in figure 3a–f, respectively. In the analysis of mass dependence of the Ljung-Box test p-values, only one test rejected the hypothesis of independence (figure 4d: k_{OLS}, $v = 2.5$ m s^{-1}) ($p = 0.2372, 0.1779, 0.6282, 0.002395, 0.279$ and 0.9537 in figure 4a–c,e,f, respectively).
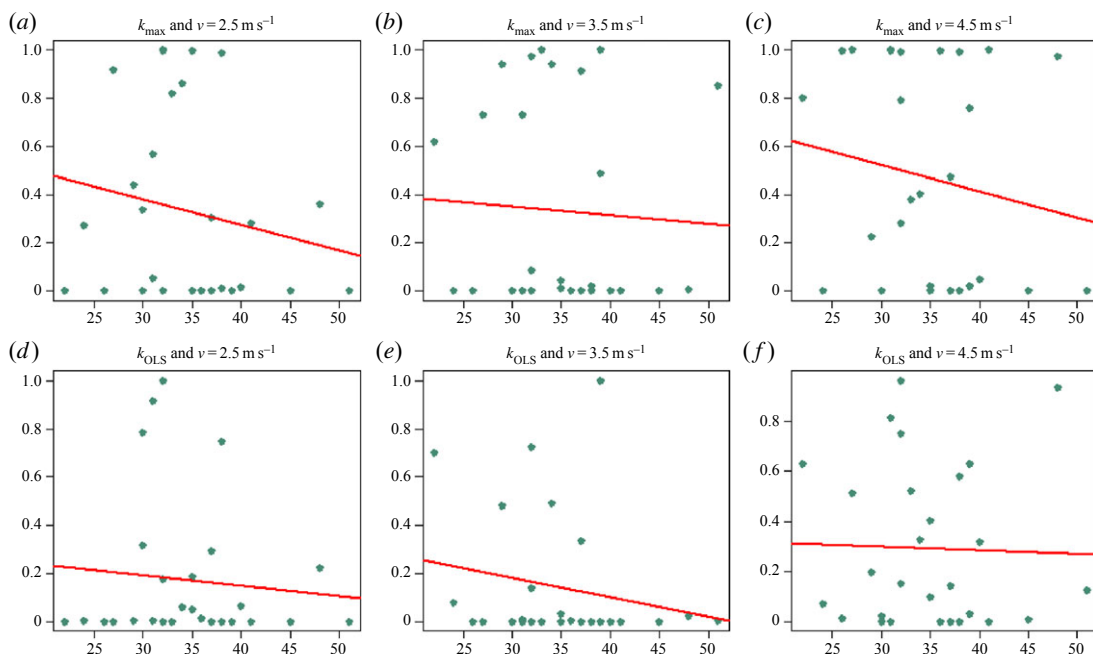
Figure 3. Dependence of the Ljung-Box test p-value with age. The y-axis represents the p-value of the Ljung-Box test at max lag 40, while the x-axis represents the age of the corresponding individuals. All of the plots show a decreasing trend but none of the relationships were statistically significant ($p = 0.3574$, 0.7737, 0.3982, 0.6224, 0.3203 and 0.8704 in (a–f), respectively).
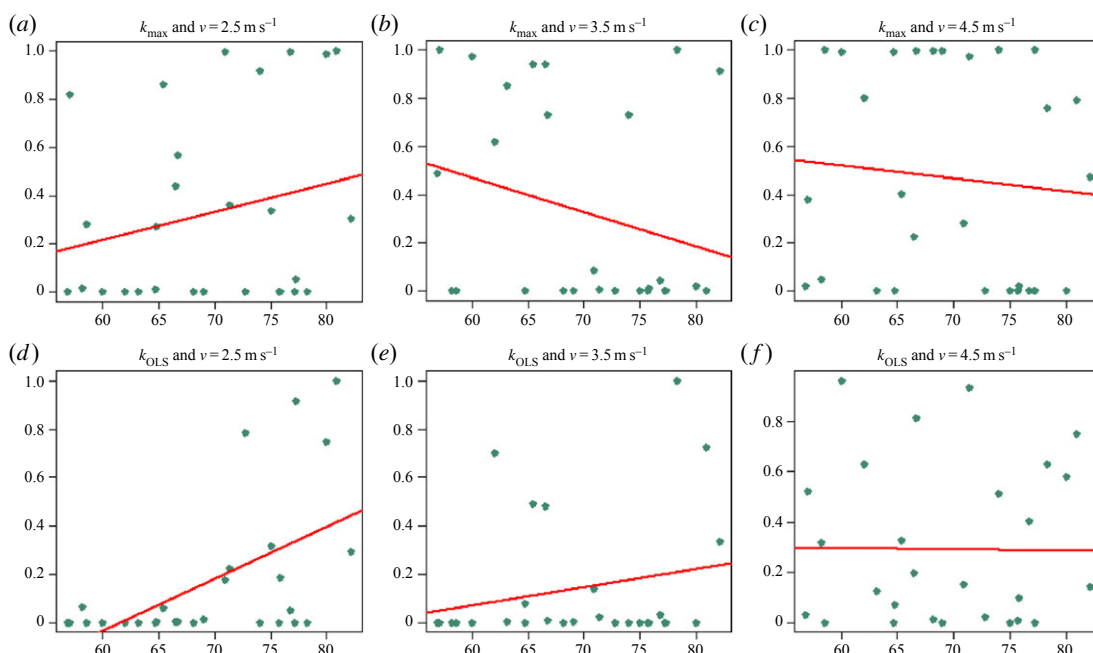

Figure 4. Dependence of the Ljung-Box test p-value with mass. The vertical axis represents the p-value of the Ljung-Box test at max lag 40, while the x-axis represents the mass of the corresponding individuals. Only one of the relationships was statistically significant ($p = 0.2372$, 0.1779, 0.6282, 0.002395, 0.279 and 0.9537 in (a–f), respectively).

3.3. Dependence of stiffness with stride and speed

The LMEMs for $k_{\max}$ and k_{OLS} with stride and velocity as fixed effects showed significant dependence of stiffness with both stride and velocity (tables 1 and 2). When we considered separate models per velocity, all models for $k_{\max}$ showed significant dependence across stride, whereas for the models for k_{OLS}, only the model for $v = 4.5 \mathrm{~m~s}^{-1}$ showed significant dependence (see tables in appendix A 4).

Table 1. Results from linear mixed effects models with k_{OLS} as the measure of stiffness. (Stride and velocity were fixed effects, and subject was a random effect. Bold indicates $p < 0.05$.)

k_{max}: linear mixed effects model					
fixed effects	estimate	s.e.	d.f.	t-value	p-value
(intercept)	15671.077	330.367	33.393	47.435	**<0.0001**
stride	−6.628	0.923	7171.831	−7.181	**<0.0001**
speed	279.633	28.404	7171.027	9.845	**<0.0001**

Table 2. Results from linear mixed effects models with k_{OLS} as the measure of stiffness. (Stride and velocity were fixed effects and subject was a random effect. Bold indicates $p < 0.05$.)

k_{OLS}: linear mixed effects model					
fixed effects	estimate	s.e.	d.f.	t-value	p-value
(intercept)	19542.831	560.299	30.892	34.879	**<0.0001**
stride	−2.850	1.258	7171.518	−2.266	**0.0235**
speed	1261.722	38.711	7171.017	32.593	**<0.0001**

4. Discussion

Our analyses demonstrate that individual strides, even when separated by several intervening strides, cannot be assumed to be independent, unless proven otherwise. Therefore, our analysis leads to the recommendation that researchers test for dependence of strides in locomotion studies with multiple strides or consider consecutive strides as dependent. Below, we discuss several aspects of this result.

4.1. Non-independence of strides

The Yule tests that we performed underlined that correlation across strides is present even at large lags (figure 1). We detected the presence of correlation of the stiffness across strides at lags over 10 for at least one of the subjects in all the scenarios that we analysed (k_{max}, k_{OLS} for $v = 2.5$ m s^{-1}, 3.5 m s^{-1}, 4.5 m s^{-1}). Given a single subject and some measurements of biomechanical parameters (e.g. CoM height, mass, age), it is not simple to determine a priori if the measurements across strides will be independent or not, so rigorous statistical hypothesis testing is needed. Dependence of locomotion behaviour across strides seems to be a diffuse phenomenon in our sample population as correlation was detected for many individuals in multiple conditions. This dependence structure of strides during the running gait was confirmed by a lag-aggregate analysis performed with the Ljung-Box test with a maximum lag of 40 (figure 2). In some of the conditions analysed, close to 50% of the individuals performed running trials with strides that were not independent. This dependence seems greater at higher velocities, which could potentially be explained by the central nervous system having insufficient time at higher velocities to respond (i.e. alter stride characteristics) fast enough to sensory feedback from environmental perturbations. In such situations, the central nervous system would rely more heavily on preflexes and on pre-determined strategies for optimal locomotion, decentralized modes, and feedforward instead of feedback control [80]. Such a scenario would result in broad similarities in the characteristics of long time series of consecutive strides. In fact, relying on centralized feedback circuits would seem to be disadvantageous because of the intrinsic time delays, especially for locomotion at high speed [80]. This explanation seems in agreement with the very fundamental structure of models such as SLIP, which show passive stabilization properties based on preflexes [80].

4.2. Stride-dependence across ages, masses, and velocities

All six conditions analysed hint at a decreasing trend of the p-value of the Ljung-Box test with age, but possess quite a high variability across individuals, especially in relative terms for those individuals with

ages between 30 and 40 years, and so no p-values were found significant at level $\alpha = 0.05$ (see figure 3). This finding is in agreement with the results of [34]. Note that it is possibly harder to observe a significant relationship between age and stiffness over the limited age range represented in this dataset. Speculatively, having a range of ages spanning life periods from childhood to older ages would help detect a stronger relationship and determine what role the interaction between reaction time and age play in the locomotor behaviour across strides. Note that we are still quite far from having an accurate neuromechanical model of human locomotion and, as far as we know, there are few results on full neuromechanical models of animal locomotion available, most of which concentrate on insects [80,81]. Speculatively, an elderly individual with respect to a younger individual running at the same speed might pay more attention at each single step and so she/he might decouple the motion of one stride from the next, which in turn would result in a lack of autocorrelation in the time series of leg stiffness or other variables calculated per stride. Given the limits of the data available (e.g. small sample size per age-range), we did not try to optimize the model and merely chose the simplest possible hypothesis class (univariate linear regression; figure 3).

Mass was not correlated with the p-value of the Ljung-Box test for any condition except one, the case of k_{OLS} and $v = 3.5 \ \mathrm{m \ s^{-1}}$ (figure 4). We do not believe the significance of this one single test is particularly meaningful and we recommend also in this case further investigation with improved sampling across a wider range of masses to better determine the relationship between mass and stride dependence during running. It is largely unknown how to describe the complex relationship between age, mass, and other variables that can come together to contribute to determining stiffness. Altogether, our analysis of the relationship between variables such as age and mass with the modulation of leg stiffness across strides while running is inconclusive. Still, it can be considered at least exploratory and makes a case for the need for further investigation into this problem.

The LMEMs determined a strong effect of both stride and velocity on leg stiffness. The models with both stride and speed as fixed effects showed statistically significant correlation at level $\alpha = 0.05$ (see tables 1 and 2). The models for each fixed velocity were all significant other than the model with k_{OLS} and $v = 2.5 \ \mathrm{m \ s^{-1}}$ (see appendix A 4 for more details about this). The dependence on velocity that we observed seem in agreement with the findings of [44], who found that the running velocity influences leg spring stiffness. However, the authors in [44] did not analyse the effect of consecutive strides on stiffness. This part of our analysis solidifies the understanding that measurements of stiffness across strides (and so strides themselves) are not independent.

4.3. Impact of ignoring dependence across strides

The variance of a sample mean from $Y_1, \ldots, Y_n$ observations, which are iid, from a population with $E[Y_1] = \mu$ and $\mathrm{Var}(Y_1) = \sigma^2$ is given by $\mathrm{Var}[\bar{Y}] = \sigma_Y^2/n$. This formula is no longer valid when there are correlations between observations. For example, in the case in which a time series is observed from a stationary stochastic process, we have:

$$\mathrm{Var}[\bar{Y}] = \frac{\sigma_Y^2}{n}\left[1 + 2\sum_{k=1}^{n}\left(1 - \frac{k}{n}\right)\rho_k\right],$$

with ρ_k the autocorrelation function at lags $k = 1, \ldots, n$. If we take the ratio between these two variances (the one for a general stationary process and the iid one), we get a quantity which contains information about the gain/loss of variance caused by neglecting possible autocorrelations in the data. This variance-to-variance ratio $R(\rho_1, \ldots, \rho_k)$ is given by

$$R(\rho_1, \ldots, \rho_k) := 1 + 2\sum_{k=1}^{n}\left(1 - \frac{k}{n}\right)\rho_k.$$

In general, positive autocorrelation is related to an increase in the variance of the mean and negative autocorrelation to a decrease in the variance of the mean. See figure 5 for some examples on how $R(\rho_1, \ldots, \rho_k)$ depends on its parameters.

Example 4.1. Suppose that $\rho_k = \rho$ for $k = 1, \ldots, n$. In this case for $\rho = 0.05$ and $n = 101$, we have $R = 6$, which means that, if we do not account for autocorrelation in the time series, we underestimate the variance of the sample mean by 6 times. Instead, if $\rho = -0.01$ and $n = 11$, then $R = 0.9$, which means that, without accounting for autocorrelation in the time series, the variance of the sample is
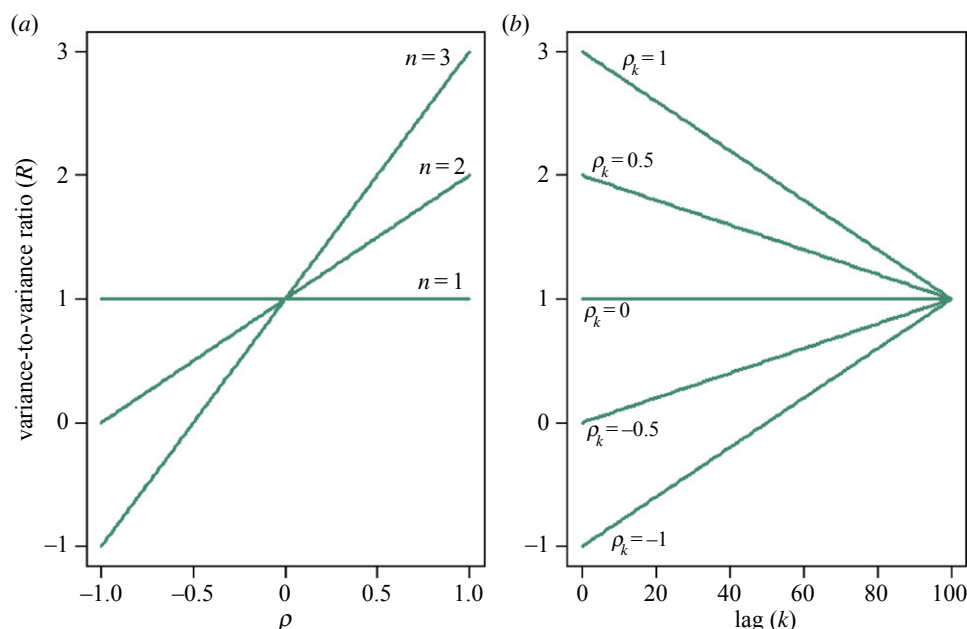
Figure 5. Variance-to-variance ratio $R(\rho_1, \ldots, \rho_k)$ with different values of the parameters ρ_k, n, and k. (a) $R(\rho, \ldots, \rho)$ for $n = 1, 2,$ 3. (b) $R(0, \ldots, 0, \rho_k, 0, \ldots, 0)$ for $k = 1, \ldots, n$ and $\rho_k = -1, -0.5, 0, 0.5, 1$.

overestimated by 10%. Note that the formula for R would not capture the full variation if $\rho_k = \rho$ for every $k \in \mathbb{N}$, as the time series would not be stationary [82].

Example 4.2. Suppose that $\rho_k = \rho$ for one single k and zero otherwise. In this case, for $\rho = 0.5$, $n = 80$ and $k = 20$, we have $R = 7/4$, which means that, if we do not account for autocorrelation in the time series, the variance of the sample mean is underestimated by 75%. Instead, if $\rho = -0.5$ and $n = 60$ and $k = 10$, then $R = 1/3$, which means that, without accounting for autocorrelation in the time series, we overestimate the variance of the sample mean by 67%.

Example 4.3. If we consider subject 2 at $v = 2.5$ m s^{-1} from our dataset and the measurements of k_{OLS} ($n = 75$ strides), the only significant autocorrelation is at lag $k = 14$ with $\rho_{14} = 0.2408617$. If we use the formula for R with only one correlation lag different from zero, we get $R = 1.391802$. Therefore, neglecting *in toto* the autocorrelation structure of the time series would fail to account for nearly 40% of the variance of the average stiffness estimate across strides, even with one single lag showing significant correlation.

Example 4.4. If we consider subject 27 at $v = 4.5$ m s^{-1} from our dataset and the measurements of k_{max} ($n = 89$ strides), and consider only the largest significant lag in the first 40 lags, namely lag $k = 8$ with $\rho_8 = 0.2111892$, we get $R = 1.384412$. Therefore, neglecting *in toto* the autocorrelation structure of the time series would again fail to account for nearly 40% of the variance of the average stiffness estimate across strides.

Example 4.5. If we consider subject 20 at $v = 2.5$ m s^{-1} from our dataset and the measurements of k_{OLS} ($n = 88$ strides), and consider all possible lags, we get $R = 0.4162655$. Therefore, neglecting *in toto* the autocorrelation structure of the time series would overestimate the variance of the average stiffness estimate across strides by nearly 60%.

Example 4.6. If we consider subject 26 at $v = 3.5$ m s^{-1} ($n = 79$ strides) from our dataset and the measurements of k_{max}, we get the coefficient $R = 0.3135405$, which implies an overestimate of the variance of the average stiffness across strides of almost 70%.

We illustrated the risk of ignoring the correlation structure in the measurements using these several examples. The most direct consequence lies in mis-evaluating the magnitude of the variability of the estimate of the average stiffness across strides. Note that in the iid case, it is well known that the variance of the sample mean decreases on the order of $1/n$, with n the sample size [67]. With correlation, things can go better or worse [82]. Positive correlation is associated with an increase in the

variance (observations tend to depart from the mean and show stochastic trend), while negative correlation is associated with a decrease in the variance (observations tend to oscillate around the mean and stay at a bounded distance from the mean over time). In our examples, we show that ignoring correlation between strides can result in either scenario. Underestimating the variance can cause spurious results (e.g. not being able to assume independence of strides), whereas overestimating the variance can prevent the detection of the fixed effects that often motivate the study in the first place (e.g. stride and/or speed association with stiffness).

One of the consequences of broad impact of the mis-estimation of the variance is the difficulty in combining the results of different laboratories. If the variance of an average measurement of a variable of interest in one experiment at one single laboratory is miscalculated, the effect could be catastrophic when results are combined across laboratories [83,84]. This is the so-called *batch effect* and it is a major and interesting problem in machine learning. Often such a problem is addressed using transfer learning [85–87], a subfield of machine learning which aims at using information extracted from one population of interest for another population of interest. Further research in this direction is required in order to mitigate possible batch effects and study replicability problems in biomechanics. These problems have not received in biomechanics the attention that they have received in other fields, such as epidemiology, public health, neuroscience and genetics [83,88,89]. More research on uncertainty quantification (e.g. estimation of the variance), together with amassing large repositories of publicly available data, could potentially facilitate the interaction of multiple laboratories and lead to faster research progress in biomechanics and locomotion.

Altogether, this analysis underlines the necessity of determining the correlation structure across strides before modelling or performing hypothesis tests. Recall that if two random variables computed by different samples are dependent, then also the samples must be dependent [67]. Therefore, the dependence of stiffness on stride indicates a dependence of the stance phases themselves within experimental trials. This is quite important information as many biomechanical studies do not perform a dependency analysis before further testing and before assuming independence. It can happen, but it is quite a 'rare' occurrence, that two random variables computed from the same sample (and that are not constructed using only one distinct element of the sample each) are independent, even if the single observations of the sample are independent. This is clearly described with rigorous mathematical theorems in [67] and it depends on a classical result in statistical inference called Basu's theorem [90]. In simple terms, the theorem states that under some regularity assumptions, minimal sufficient statistics for a parameter are independent of any ancillary statistics for that parameter. For example, the normal distribution is characterized by the condition that the sample mean (which is minimal sufficient for the population mean) is independent of the sample variance (which is ancillary for the population mean). With the word 'characterized' we mean that no other distribution possesses this property.

4.4. Overground experiments and the case of few strides per bout

In our analysis, we discussed to what extent autocorrelation between consecutive strides within a given trial of an experiment is present. We did not discuss extensively what happens to the variability of a measurement like leg stiffness in the case a researcher decides to estimate the parameter using a single stride per bout and to use multiple bouts. Using strides from different trials might require assessments such as: which stride of each bout shall we select? The one in the middle? Why? Is there a way to fairly compare bouts that have different numbers of strides? We believe that the answers to these questions require further data-driven research. Still, the strategy of using only one stride per bout risks to be sub-optimal because it does not seem reasonable to decide *a priori* to throw away a significant portion of the data merely because there might be a statistically yet-to-be-confirmed dependency across strides or because one does not want to deal with the dependency structure. Even just considering the experimental effort and the cost in terms of time, it is important to get the most out of the data collected. As shown in this manuscript, the statistical tools to harness stride dependency and to give a correct uncertainty quantification of the estimate of the parameter of interest are available. We recommend assessing the dependency structure of the sample also in the case in which an investigator decides to choose one single stride per bout and consider multiple bouts, but also in every other estimation method used. Again, failing to correctly assess the variability of a measurement (e.g. stiffness) might result in spurious findings or in losing the possibility to uncover important biological information. The determination of what the greatest sources of

dependency (e.g. variation across strides versus variation across trials) are can have an important impact on the design of locomotion experiments and surely deserves further attention.

If multiple strides are available, it is really valuable information and, unless there has been some experimental failure owing to which some strides need to be excluded, it is important to keep them and evaluate them. We potentially understand the use of a single stride in the case in which the study is concentrated on that single stride, but even in such a case there is something to say. Suppose that a researcher wants to understand the behaviour of individuals during the first stride of a locomotor trial. Then, there is one single first stride in each experiment. Especially if the strides are independent, it is a reasonable choice to use only the first stride of the experiment for any estimate related to the first stride. However, the situation here is also tricky because if there is autocorrelation at lag 1, some information about the first stride is shadowed in the second stride, too. In a similar way, there is potentially some information about the first stride in other strides as well. Note that autocorrelation at lag 1 does not mean that stride 1 is correlated to stride 2 or that stride 2 is correlated to stride 3 on an individual stride-by-stride basis. Rather, it means that there is correlation between consecutive strides globally in the data. In the simplest autoregressive model AR(1) [82], there are two parameters: the variance of the noise and the correlation coefficient ρ between one observation and the next. In such a case, $\rho_k = \rho^k$, with k the lag considered. This means that the correlation predicted by the AR(1) model decays exponentially with the lag, but it is present at every lag. If, for example, the correlation is 0.7 at lag 1, at lag 2 the correlation is 0.49. The implication of this is that, removing only a bunch of strides from the sample might not remove in full the autocorrelation among the remaining strides. The case of correlation at large lags is a statistically more interesting case, but also very interesting from the biomechanical perspective because, although one might intuitively expect a relationship from one stride to the next, one might not expect to see a relationship between one stride and strides that take place 5 to 10 strides later.

It is more common for data containing large numbers of strides to be available during running on a treadmill than overground or on uneven ground [48]. Therefore, the implications of the present study may be most relevant to treadmill-based locomotion studies. In overground studies with stationary trackways and force plates, the animal moves across the field of view just once and so the number of strides per bout available for analysis is more limited. The dataset that we analysed [28] is of treadmill running and so we concentrated more on uncertainty quantification in that type of experiment. However, parts of our discussion also apply to the case in which fewer strides are available. In fact, although we highlight in figure 1 those subjects whose strides showed autocorrelation at large lags (we put a threshold at lag 5), figure 1 shows that autocorrelation was also present at lower lags.

If relatively few strides are available (e.g. overground running), then the variance can be computed only using those strides, with calculations analogous to those in the examples of §4.3 (e.g. example 2, but with a smaller lag). Still, if there is dependence between one stride and another, many strides can provide information about many others. As noted in our analysis, some strides might be outliers. If it is determined that a stride is an outlier via a careful verification with a test for outliers, the exclusion of that stride is recommended. Care must be taken, especially in the case in which only a few strides are available. A popular test for outliers is Grubbs's test [91], but such a test gives reasonable results only when there are enough observations, as makes intuitive sense. In the case in which too few strides are available, Grubbs's test will, most likely, overcount outliers. It is a general fact that when the sample is small, the estimation of the variance has limited validity. The extreme case of considering one stride per trial is essentially kind of equivalent to implicitly assuming that there is no variability across strides. In this paper, we provided evidence that such an assumption might be invalid. When it is not necessary to consider only one single stride per trial, such a decision should not be made lightly because there is no physical principle or mathematical law that explains the relationship between the value of many locomotor parameters like stiffness and the stride number in a locomotor bout. The problem of correctly extracting information about the variability of an estimate in overground locomotion may be first experimental, and relate to data collection, before potentially being statistical. In the case of human locomotion, until the instrumentation to record more than a few dozen metres of overground running is widely available, then researchers will always have relatively few strides available for study and comparatively little information about uncertainty quantification for parameters of interest. In turn, this will negatively impact the replicability of overground studies across laboratories.

4.5. Other considerations

For what concerns the estimation of stiffness *in se*, there is always the question: what stiffness are we really estimating? [42,43]. As far as we know, there is no mathematical model that describes in a

satisfactorily quantitative way the process through which the brain modulates leg stiffness. Reductionist models, such as SLIP, describe the most essential locomotor dynamics with the most parsimonious model in terms of the number of variables used [29,30,80]. In our *model 1—k_{max}*, we used a method which is biologically and mechanically well-motivated by the very definition of spring constant and elastic force. In our *model 2—k_{OLS}*, we considered the stiffness as the least square estimates of a simple linear regression model without intercept [92]. Note that this second method might be criticizable for the possibly more cumbersome biological interpretation. However, the first method, as well, is questionable from the biological perspective, as the stiffness is not associated with the stiffness of any specific tissue, but it is a resulting parameter of a complex system which collapsed from a multi-legged body with distributed mass to a single point mass on a spring [80]. We believe that k_{max} and k_{OLS} are good examples for the purposes of our paper, as they emerge from quite alternative perspectives (biologically driven k_{max} and data-driven k_{OLS}). Incorporating a broad comparison of all possible methods for calculating stiffness is beyond the scope of this paper. Such an analysis would possibly encounter a similar across-strides-dependence problem and possibly lead to analogous results and consequences.

Note that for several of the subjects, there is a possible outlier in the stiffness measurements of the first strides (figures 7–12). We did not remove those observations before fitting the model, as it is actually relevant to know that, in some subjects, the first stride needs to be tested as a possible outlier. As mentioned in the introduction, biomechanists often address this first-stride-problem by removing the first and last strides of an experiment from the analysed dataset [1]. Sometimes those strides are not even recorded because they occur off camera. Again, we consider removal strategies suboptimal; in this case, removing initial and final strides compromises the opportunity to model the initial transition phase. Depending on the study, researchers might not be interested in the opportunity of modelling transitional phases. In such cases, first and last strides are removed from consideration because they might contain a noisy signal. Note that, also in such cases, careful statistical considerations must be made before deciding on the removal of those strides. How many strides should be removed? This is a tricky question. In some cases, after the removal of an outlier with the support of a rigorous statistical test, the same test (e.g. Grubbs's test [91]) on the reduced dataset (without the observation that was determined to be an outlier after the first test) might detect an additional outlier that was not detected when the test was initially performed on the full dataset. Therefore, we recommend to thoughtfully test for outliers also in studies of steady-state locomotion, before making the decision to remove some strides from the data, and to quantitatively support the decision to consider some strides transitional. If the first or last strides are not outliers, there is not really a statistical justification in support of removing such observations from the analysis. Although the removal of first and last strides is perceived as safe and convincing from a biomechanical perspective, if those strides are not outliers, curtailing the data in this way will end up having a negative impact on the correct estimation of variability, with the worst effects occurring when the number of available strides is lowest (e.g. experiments taking place overground or on uneven ground).

5. Conclusion

As far as we know, this is the first paper in the terrestrial biomechanics literature dedicated to discussing rigorously and extensively the effects of neglecting the correlation structure of consecutive strides. We found evidence of autocorrelation of the time series of human leg stiffness across strides during running on a treadmill and, thereby, evidence of the dependence of the time series of kinematic and kinetic data at different strides. Therefore, our paper has implications for the diffuse practice in biomechanics of assuming that individual strides within an experimental trial are independent and identically distributed. Lacking an understanding of stride dependence structure invalidates estimates of uncertainty of average measurements across strides, including the variance. Variability assessments are required for both building correct confidence intervals and valid hypothesis testing. The incorrect quantification of uncertainty might inflate biological signals in some cases or miss it in others. We caution against the practice of blindly assuming stride independence and recommend that researchers first test for stride independence before assuming it. If, after testing, the strides turn out to not be independent, we recommend exploiting the correlation structure. By contrast, retaining only one of every few strides is a suboptimal remedy because it decreases the sample size by an order of magnitude and risks to disrupt the correlation structure of the time series collected from the experiment. Having an incorrect estimate of the variability of biologically important quantities such as leg stiffness or any other measurement can negatively impact the replicability of the study. We argue that it is important not only to include measurements of uncertainty peculiar to an

experiment in every study, but also to have a standardized methodology to assess uncertainty. Not only would this facilitate replicability and comparison, but it would also foster collaboration across research groups.

Appendix A

This appendix contains results of some additional analyses that we performed.

A.1. Stiffness trajectories at different speeds

Figure 6 below shows the different trajectories of k_{OLS} and k_{max} for all the recorded stride of every subject for the three speeds considered $v = 2.5$ m s^{-1}, 3.5 m s^{-1} and 4.5 m s^{-1}.
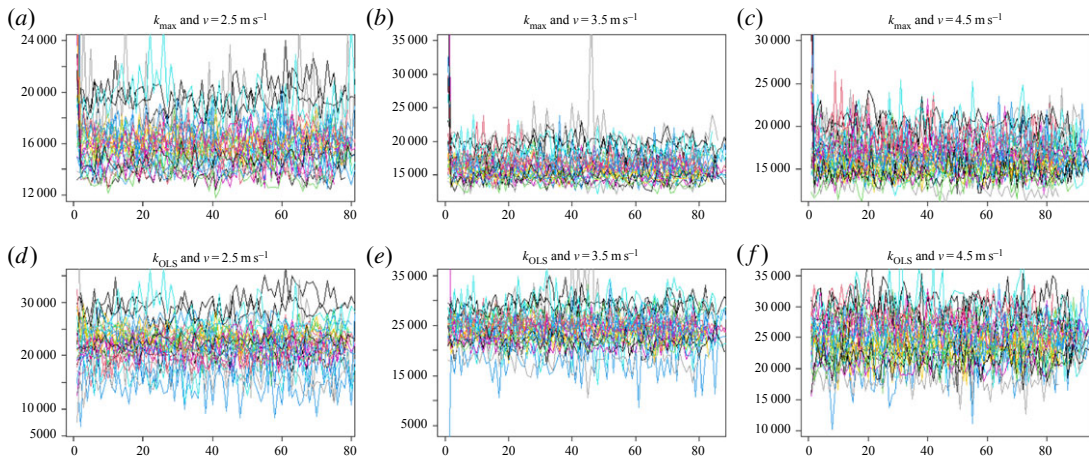

Figure 6. Trajectories of k_{max} (a–c) and k_{OLS} (d–f) of each subject across the three speeds considered $v = 2.5$ m s^{-1} (a,d), 3.5 m s^{-1} (b,e) and 4.5 m s^{-1} (c,f) for all the recorded strides. The y-axes represent stiffness k and the x-axes represent the stride number.

A.2. Deterministic trend of stiffness across strides per subject

In this section, we checked for the presence of a deterministic trend of human leg stiffness across strides using both the methods for calculating stiffness (k_{OLS} and k_{max}). The models we fitted were

$$k_i = \beta_0 + \beta_1 S_i + \epsilon_i, \quad i = 1, \ldots, 28,$$

with k_i being k_{OLS} and k_{max} and S_i, $i = 1, \ldots, 28$ being the stride number. In total, we had six of these models, one for each combination of velocity ($v = 2.5$ m s^{-1}, $v = 3.5$ m s^{-1} and $v = 4.5$ m s^{-1}) and stiffness k ($k = k_{OLS}$ and $k = k_{max}$). The level of significance was $\alpha = 0.05$ and p-values were calculated with 4 decimals of significance. In some of the subjects and for some of the conditions and

measurements of stiffness, there was what appeared to be an outlier at the beginning of the stride considered. As mentioned in the core part of this manuscript (§4.5), we decided to keep that first observation as it was significant for our discussion, given that its presence further supports the need to consider dependence and variability across strides.

The analysis of the deterministic trend, always modelled as linear in our models, showed some significant results across conditions and subjects for both stiffness measurements k_{OLS}, k_{max}. The significance of this result is not that we determined that the stiffness changes across stride linearly, but that there might be such a change, and it is risky to neglect such a relationship.
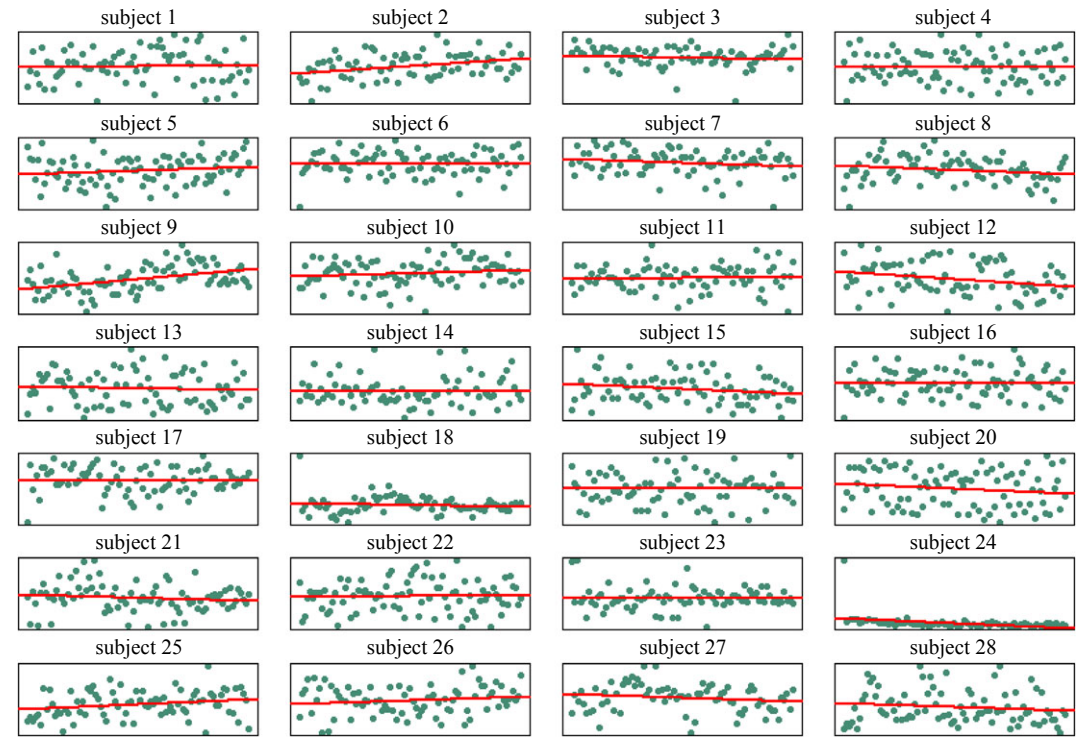

Figure 7. Deterministic trend of k_{OLS} across strides for each of the 28 subjects at $v = 2.5$ m s^{-1}.

k_{OLS} and $v = 2.5$ m s^{-1}

Table 3. The p-values produced when testing for the presence of a significant deterministic linear relationship between the stride number and k_{OLS} for each of the 28 subjects at $v = 2.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

subject 1	subject 2	subject 3	subject 4
0.8411	**0.0041**	0.3339	0.7724
subject 5	subject 6	subject 7	subject 8
0.1727	0.6668	0.1108	0.1182
subject 9	subject 10	subject 11	subject 12
0.0001	0.1243	0.7434	**0.0163**
subject 13	subject 14	subject 15	subject 16
0.5029	0.7842	0.0688	0.9185
subject 17	subject 18	subject 19	subject 20
0.9087	0.4223	0.9920	0.2044
subject 21	subject 22	subject 23	subject 24
0.2409	0.6083	0.7839	**0.0025**
subject 25	subject 26	subject 27	subject 28
0.0382	0.2333	0.1686	0.2378

There is a significant correlation between stride S and stiffness k_{OLS} in 5 out of 28 subjects in the case $v = 2.5\,\mathrm{m\,s}^{-1}$ (figure 7 and table 3).

Figure 8. Deterministic trend of k_{OLS} across strides for each of the 28 subjects at $v = 3.5\,\mathrm{m\,s}^{-1}$.

$\underline{k_{\mathrm{OLS}}\ \text{and}\ v = 3.5\ \mathrm{m\,s}^{-1}}$

Table 4. The p-values produced when testing for the presence of a significant deterministic linear relationship between the stride number and k_{OLS} for each of the 28 subjects at $v = 3.5\,\mathrm{m\,s}^{-1}$. (Bold indicates $p < 0.05$.)

subject 1	subject 2	subject 3	subject 4
0.8348	0.0560	0.2325	0.3368
subject 5	subject 6	subject 7	subject 8
0.4128	0.2853	bf 0.0168	0.0581
subject 9	subject 10	subject 11	subject 12
0.6165	0.9244	0.9203	0.7140
subject 13	subject 14	subject 15	subject 16
<0.0001	0.2614	0.3294	0.6886
subject 17	subject 18	subject 19	subject 20
0.7568	0.7385	**0.0002**	0.6169
subject 21	subject 22	subject 23	subject 24
0.2495	0.4178	0.7642	0.3234
subject 25	subject 26	subject 27	subject 28
0.1185	0.2968	**0.0483**	0.3173

There is a significant correlation between stride S and stiffness k_{OLS} in 4 out of 28 subjects in the case $v = 3.5\,\mathrm{m\,s}^{-1}$ (figure 8 and table 4).

Figure 9. Deterministic trend of k_{OLS} across strides for each of the 28 subjects at $v = 4.5 \ \mathrm{m\ s^{-1}}$.

k_{OLS} and $v = 4.5 \ \mathrm{m\ s^{-1}}$

Table 5. The p-values produced when testing for the presence of a significant deterministic linear relationship between the stride number and k_{OLS} for each of the 28 subjects at $v = 4.5 \ \mathrm{m\ s^{-1}}$. (Bold indicates $p < 0.05$.)

subject 1	subject 2	subject 3	subject 4
0.8411	**0.0041**	0.3339	0.7724
subject 5	subject 6	subject 7	subject 8
0.1727	0.6668	0.1108	0.1182
subject 9	subject 10	subject 11	subject 12
0.0001	0.1243	0.7434	**0.0163**
subject 13	subject 14	subject 15	subject 16
0.5029	0.7842	0.0688	0.9185
subject 17	subject 18	subject 19	subject 20
0.9087	0.4223	0.9920	0.2044
subject 21	subject 22	subject 23	subject 24
0.2409	0.6083	0.7839	**0.0025**
subject 25	subject 26	subject 27	subject 28
0.0382	0.2333	0.1686	0.2378

There is a significant correlation between stride S and stiffness k_{OLS} in 5 out of 28 subjects in the case $v = 4.5 \ \mathrm{m\ s^{-1}}$ (figure 9 and table 5).

Figure 10. Deterministic trend of k_{max} across strides for each of the 28 subjects at $v = 2.5$ m s^{-1}.

$\underline{k_{max} \text{ and } v = 2.5 \text{ m s}^{-1}}$

Table 6. The p-values produced when testing for the presence of a significant deterministic linear relationship between the stride number and k_{max} for each of the 28 subjects at $v = 2.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

subject 1	subject 2	subject 3	subject 4
0.4520	0.6103	**0.0162**	0.1165
subject 5	subject 6	subject 7	subject 8
0.2355	0.4618	0.9171	0.4191
subject 9	subject 10	subject 11	subject 12
0.0184	0.3564	0.4658	0.4379
subject 13	subject 14	subject 15	ssubject 16
0.6891	0.9220	**0.0176**	0.6669
subject 17	subject 18	subject 19	subject 20
0.2488	0.2617	0.4096	0.1655
subject 21	subject 22	subject 23	subject 24
0.1410	0.4899	0.0862	**0.0083**
subject 25	subject 26	subject 27	subject 28
0.8918	0.2867	**0.0119**	0.1215

There is a significant correlation between stride S and stiffness k_{max} in 5 out of 28 subjects in the case $v = 2.5$ m s^{-1} (figure 10 and table 6).
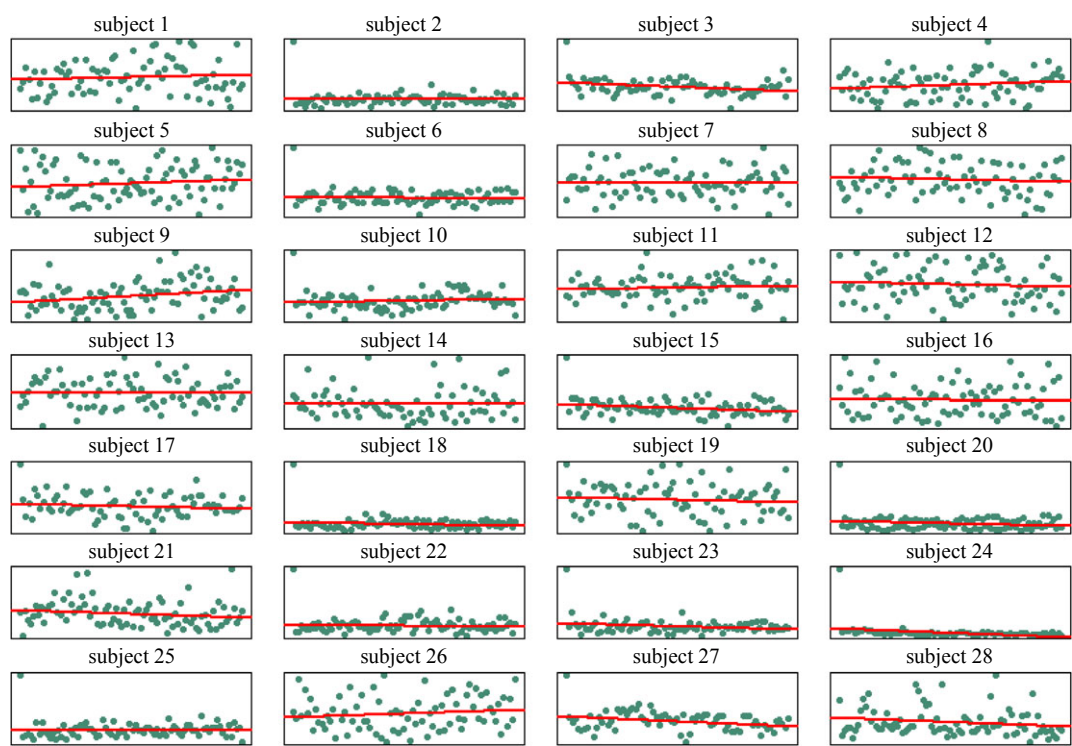
Figure 11. Deterministic trend of k_{max} across strides for each of the 28 subjects at $v = 3.5$ m s^{-1}.

$\underline{k_{max} \text{ and } v = 3.5 \text{ m s}^{-1}}$

Table 7. The p-values produced when testing for the presence of a significant deterministic linear relationship between the stride number S and k_{max} for each of the 28 subjects at $v = 3.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

subject 1	subject 2	subject 3	subject 4
0.2106	**0.0060**	0.1172	0.1994
subject 5	subject 6	subject 7	subject 8
0.0829	0.2021	**0.0012**	0.5693
subject 9	subject 10	subject 11	subject 12
0.4309	0.6504	0.7665	0.6102
subject 13	subject 14	subject 15	subject 16
<0.0001	0.1913	0.3728	0.6726
subject 17	subject 18	subject 19	subject 20
0.8171	0.2153	**0.0031**	0.7889
subject 21	subject 22	subject 23	subject 24
0.5151	0.7966	0.2783	0.2989
subject 25	subject 26	subject 27	subject 28
0.1207	0.0742	0.1075	0.1732

There is a significant correlation between stride S and stiffness k_{max} in 4 out of 28 subjects in the case $v = 3.5$ m s^{-1} (figure 11 and table 7).
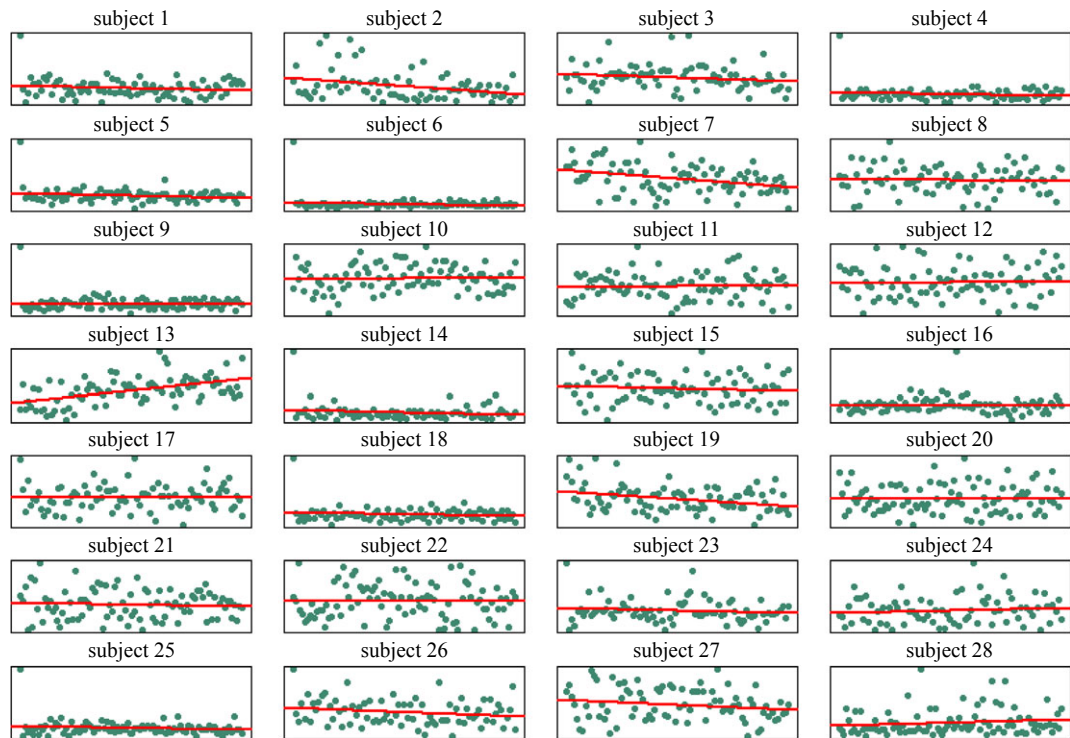
Figure 12. Deterministic trend of k_{max} across strides for each of the 28 subjects at $v = 4.5$ m s^{-1}.

$\underline{k_{max} \text{ and } v = 4.5 \text{ m s}^{-1}}$

Table 8. The p-values produced when testing for the presence of a significant deterministic linear relationship between the stride number S and k_{max} for each of the 28 subjects at $v = 4.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

subject 1	subject 2	subject 3	subject 4
0.0601	0.4452	0.4808	0.5843
subject 5	subject 6	subject 7	subject 8
0.7060	0.0906	0.1709	**0.0005**
subject 9	subject 10	subject 11	subject 12
0.1151	**0.0253**	0.2640	0.9275
subject 13	subject 14	subject 15	subject 16
0.0989	0.3569	0.0777	0.5435
subject 17	subject 18	subject 19	subject 20
0.2608	**0.0045**	0.7787	0.1241
subject 21	subject 22	subject 23	subject 24
0.2418	**0.0042**	0.1774	0.3052
subject 25	subject 26	subject 27	subject 28
0.2701	**0.0042**	0.2739	0.1942

There is a significant correlation between stride S and stiffness k_{max} in 5 out of 28 subjects in the case $v = 4.5$ m s^{-1} (figure 12 and table 8).
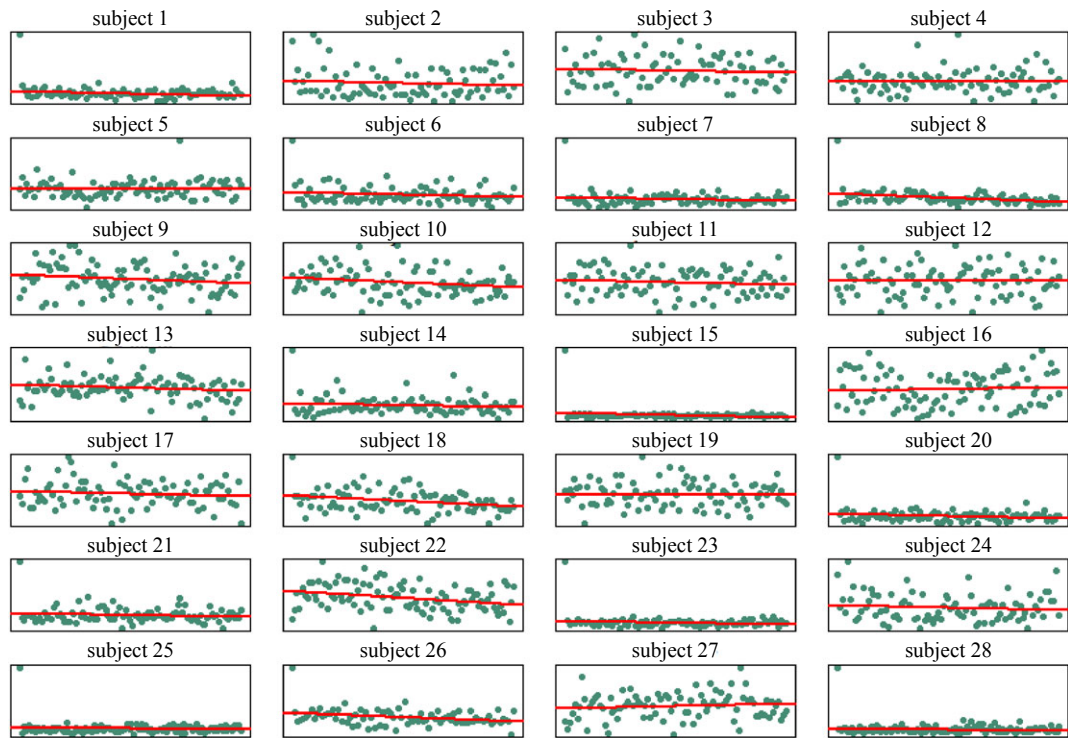
A.3. Autocorrelation of stiffness across strides per subject

In this subsection, we computed and plotted the sample ACF $\hat{\rho}_Y(k) := \hat{\gamma}_Y(k)/\hat{\gamma}_Y(0)$ to estimate the ACF $\rho_Y(k) := \text{Corr}(Y_t, Y_{t+k})$ for $k = 1, \ldots, 40$ and both measures of stiffness (k_{OLS} and k_{max}) for each subject $i = 1, \ldots, 28$ and velocity $v = 2.5$ m s^{-1}, $v = 3.5$ m s^{-1} and $v = 4.5$ m s^{-1}. Our main goal was to detect if the observations k_{OLS} and k_{max}, measured at different strides, are iid or not. When k starts to be only slightly smaller than n, the estimates $\hat{\rho}_Y(k)$, $\hat{\gamma}_Y(k)$ are unreliable since there are only a few observations

to use for the estimates. Therefore, we stopped the lag at 40, which was reasonably less than the minimum (across subjects) number of strides observed.

Note that $\rho_Y(k)$ gives the full description of the covariance structure of the time series Y_t only in the stationary case, but not in the non-stationary case. In the non-stationary case, the ACF depends on both the lag and the position of the random variables in the time-series order [82]. We do not enter here in the full analysis of what more can go wrong in the case in which the time series of strides are non-stationary (to keep the paper of a reasonable length, we did not include this analysis for our data). In the non-stationary case, the iid hypothesis can be violated also in the case in which observations are independent (and so they must be not identically distributed). This happens, for example, for a time series $Y_t \sim \mathcal{N}(0, \exp(t))$ with Y_{t_1} independent of Y_{t_2} for every $t_1 \neq t_2$. Note that, in this case, the variance is growing exponentially in time, and so, naturally, later observations are more dispersed than earlier observations. This needs to be taken into account for an accurate quantification of uncertainty. Recall, that a way to understand if a time series is not stationary is by checking if $\rho_Y(k) \to 0$. Indeed, it is well known (See Proposition 2.2 [82]) that for causal ARMA processes, $\rho_Y(k) \to 0$ at an exponential rate as $|k| \to +\infty$. By Wold decomposition theorem, a purely non-deterministic Gaussian process is an $MA(\infty)$ process with normal white noise (See Proposition 2.2 [82]). This argument tells us that a lack of fast decay in the ACF, is a hint of non-stationarity of the process. The lack of decay to zero of the ACF implies an even further distance from zero of the ACF. Therefore, non-stationarity implies an even stronger deviation from the iid hypothesis [82] and so a further complication for the analysis of uncertainty. As it is depicted in figures 13–18, the analysis of the autocorrelation function through strides for k_{OLS} and k_{max} for each subject at all velocity conditions $v = 2.5 \text{ m s}^{-1}$, $v = 3.5 \text{ m s}^{-1}$ and $v = 4.5 \text{ m s}^{-1}$ indicates that, potentially, some of the time series were non-stationary [82].

k_{OLS} and $v = 2.5 \text{ m s}^{-1}$ The ACF of 9 out of 28 subjects does not seem to decay, indicating that the time series might be non-stationary.
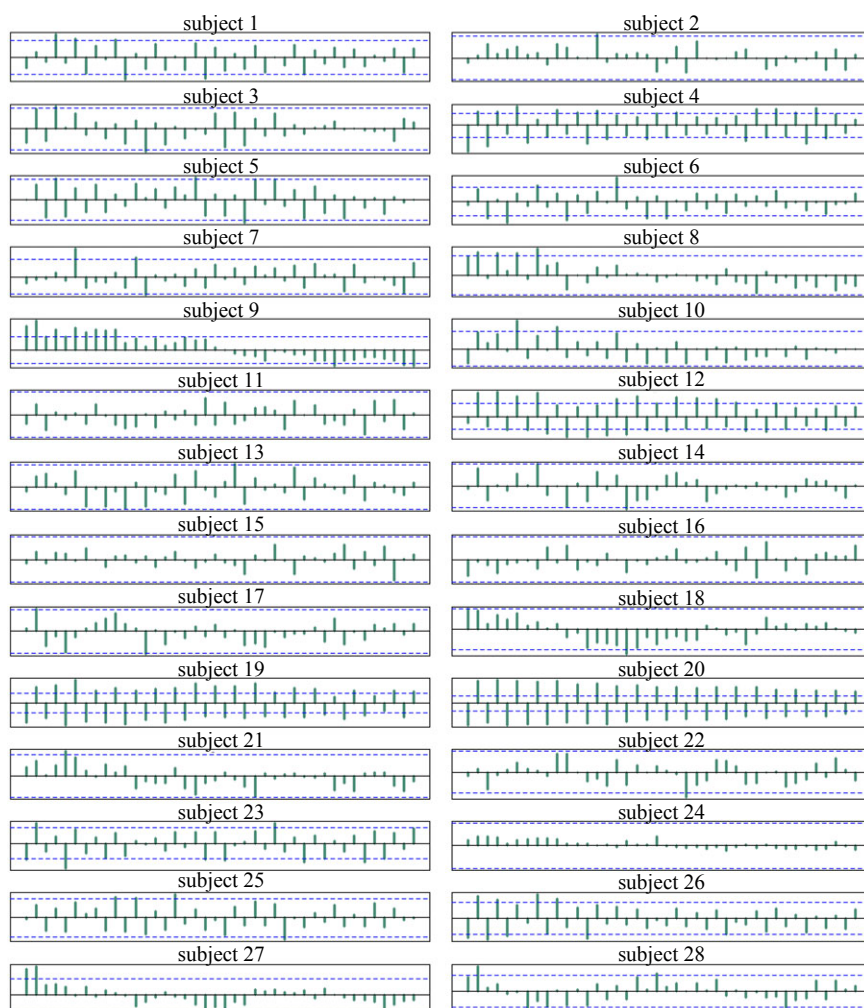

Figure 13. Autocorrelation function of k_{OLS} across strides for each of the 28 subjects at $v = 2.5 \text{ m s}^{-1}$. When a green vertical bar exceeding the blue dotted line in either positive or negative directions, it indicates the presence of significant autocorrelation at that lag.

k_{OLS} and $v = 3.5 \mathrm{~m~s}^{-1}$ The ACF of 9 out of 28 subjects does not seem to decay, indicating that the time series might be non-stationary.

Figure 14. Autocorrelation function of k_{OLS} across strides for each of the 28 subjects at $v = 3.5 \mathrm{~m~s}^{-1}$. When a green vertical bar exceeding the blue dotted line in either positive or negative directions, it indicates the presence of significant autocorrelation at that lag.

k_{OLS} and $v = 4.5 \text{ m s}^{-1}$ The ACF of 5 out of 28 subjects does not seem to decay, indicating that the time series might be non-stationary.

Figure 15. Autocorrelation function of k_{OLS} across strides for each of the 28 subjects at $v = 4.5 \text{ m s}^{-1}$. When a green vertical bar exceeding the blue dotted line in either positive or negative directions, it indicates the presence of significant autocorrelation at that lag.
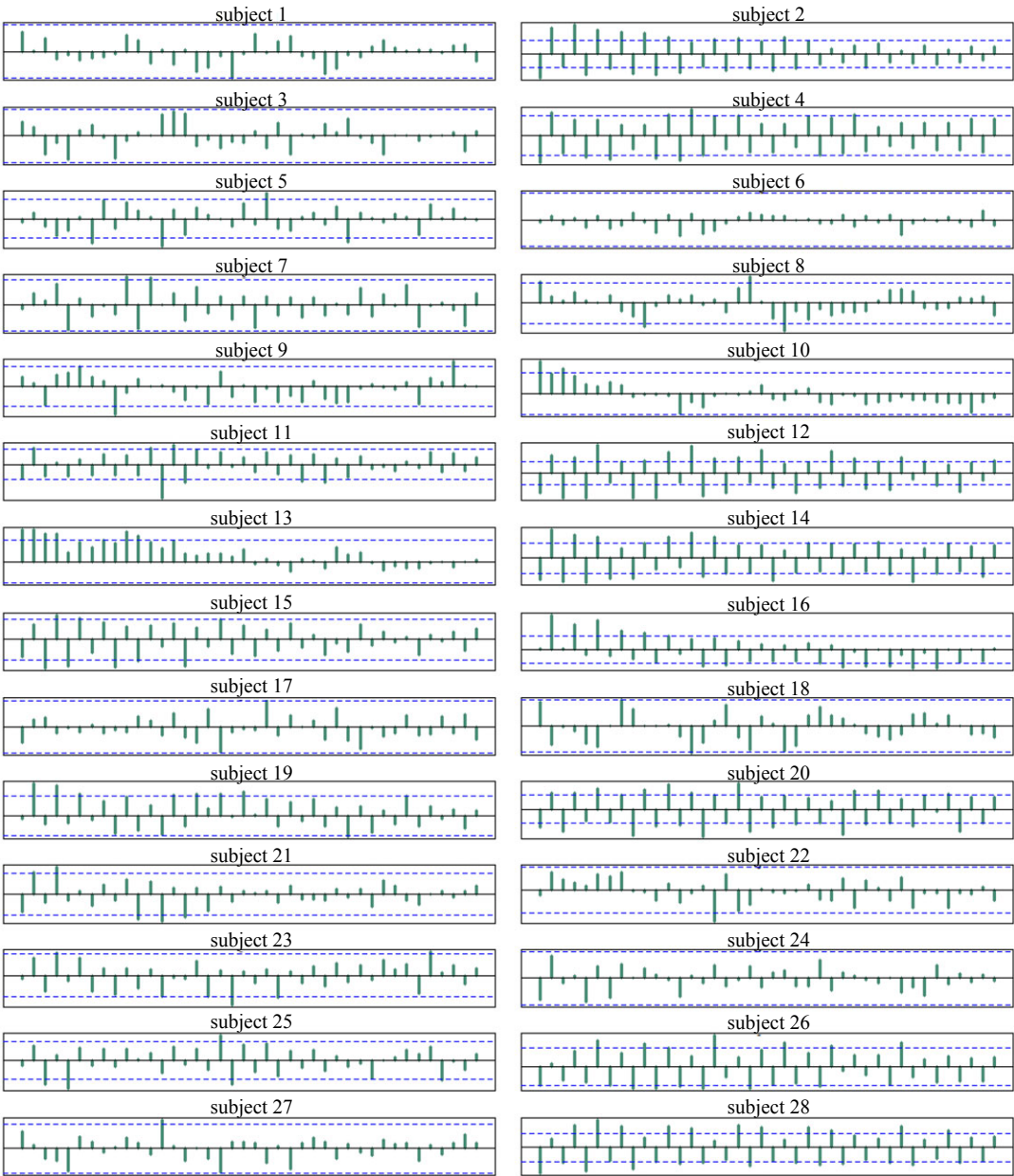
k_{max} and $v = 2.5$ m s^{-1} The ACF of 7 out of 28 subjects does not seem to decay, indicating that the time series might be non-stationary.

Figure 16. Autocorrelation function of k_{max} across strides for each of the 28 subjects at $v = 2.5$ m s^{-1}. When a green vertical bar exceeding the blue dotted line in either positive or negative directions, it indicates the presence of significant autocorrelation at that lag.

$k_{\max}$ and $v = 3.5 \text{ m s}^{-1}$ The ACF of 9 out of 28 subjects does not seem to decay, indicating that the time series might be non-stationary.
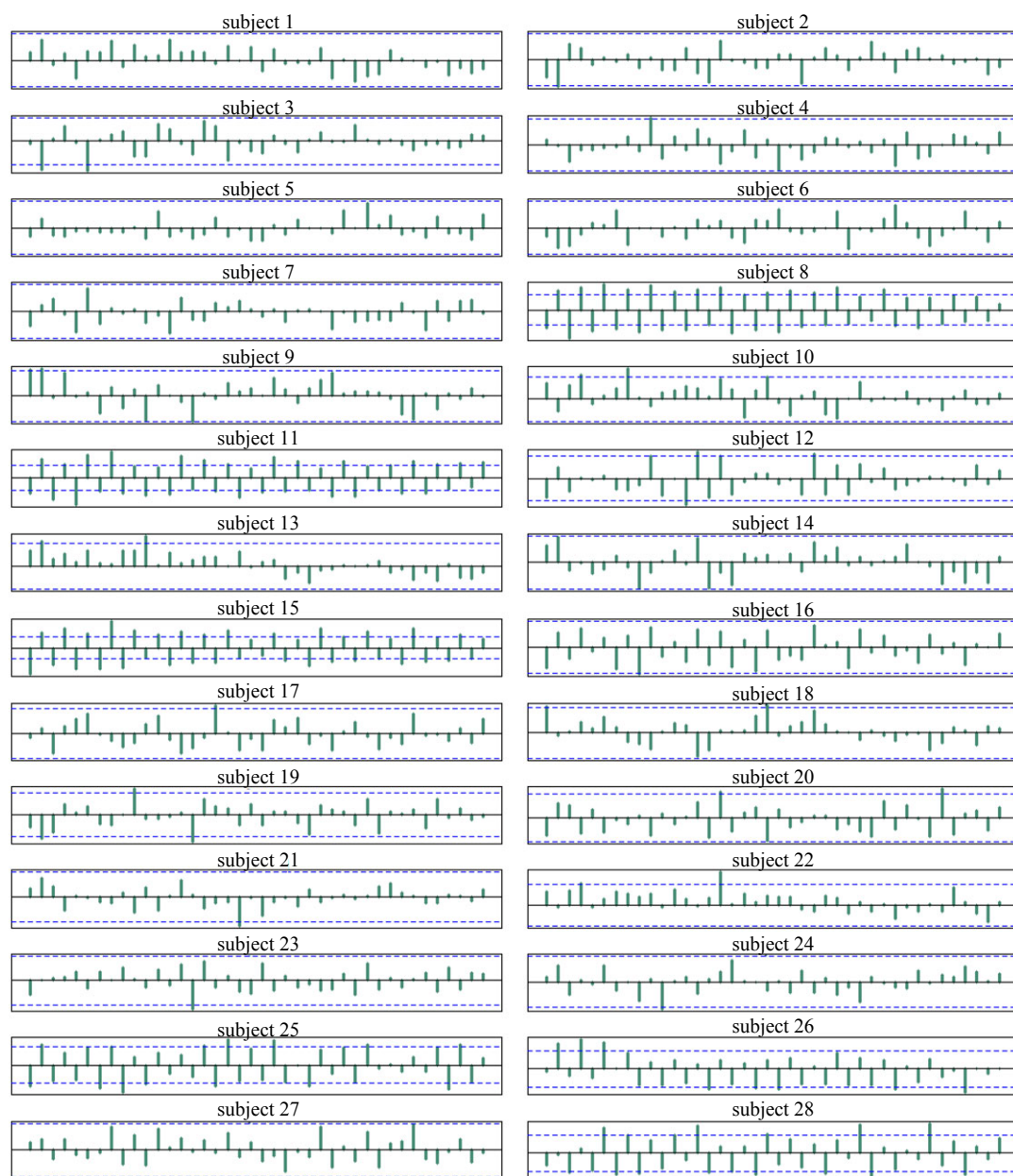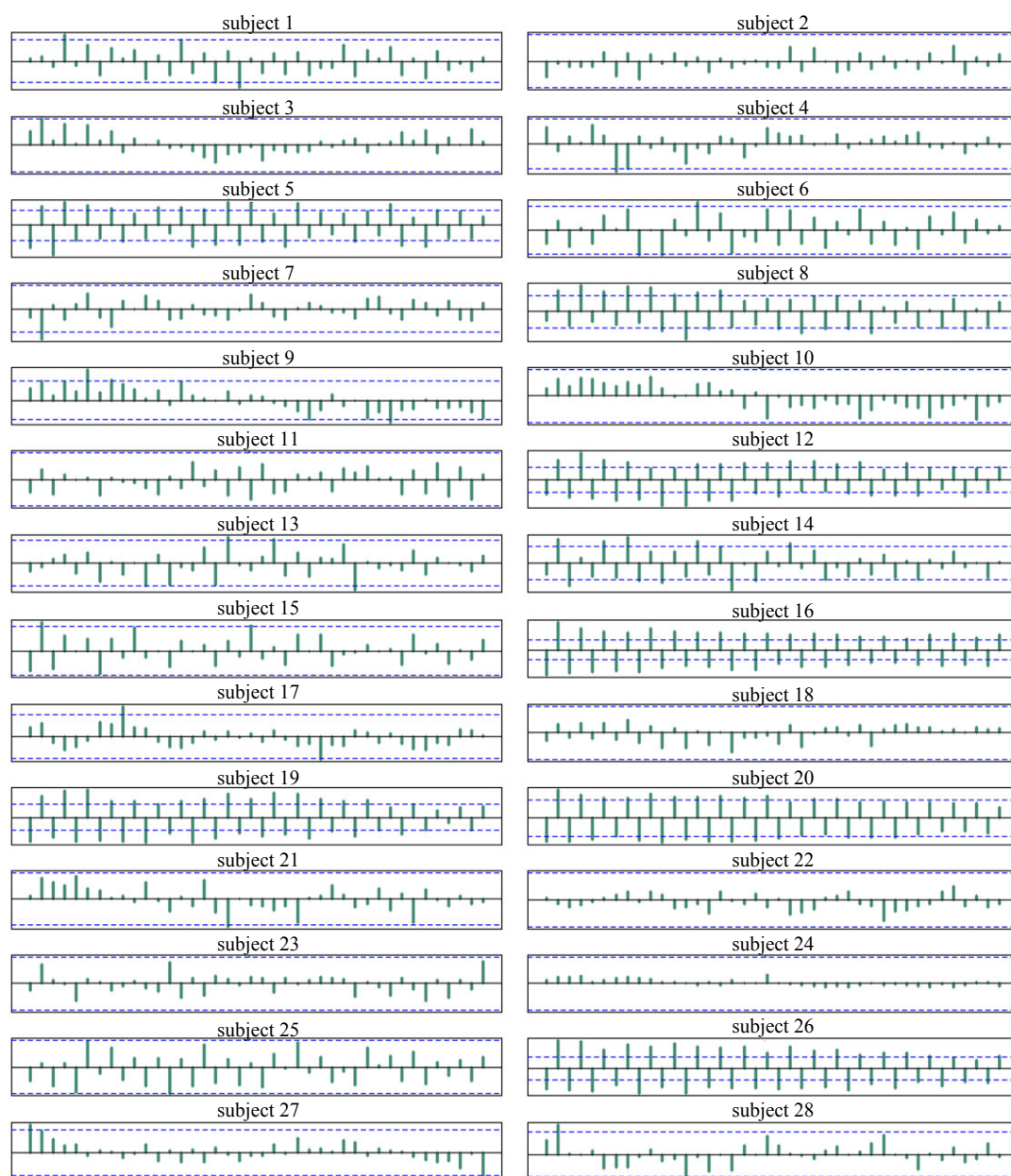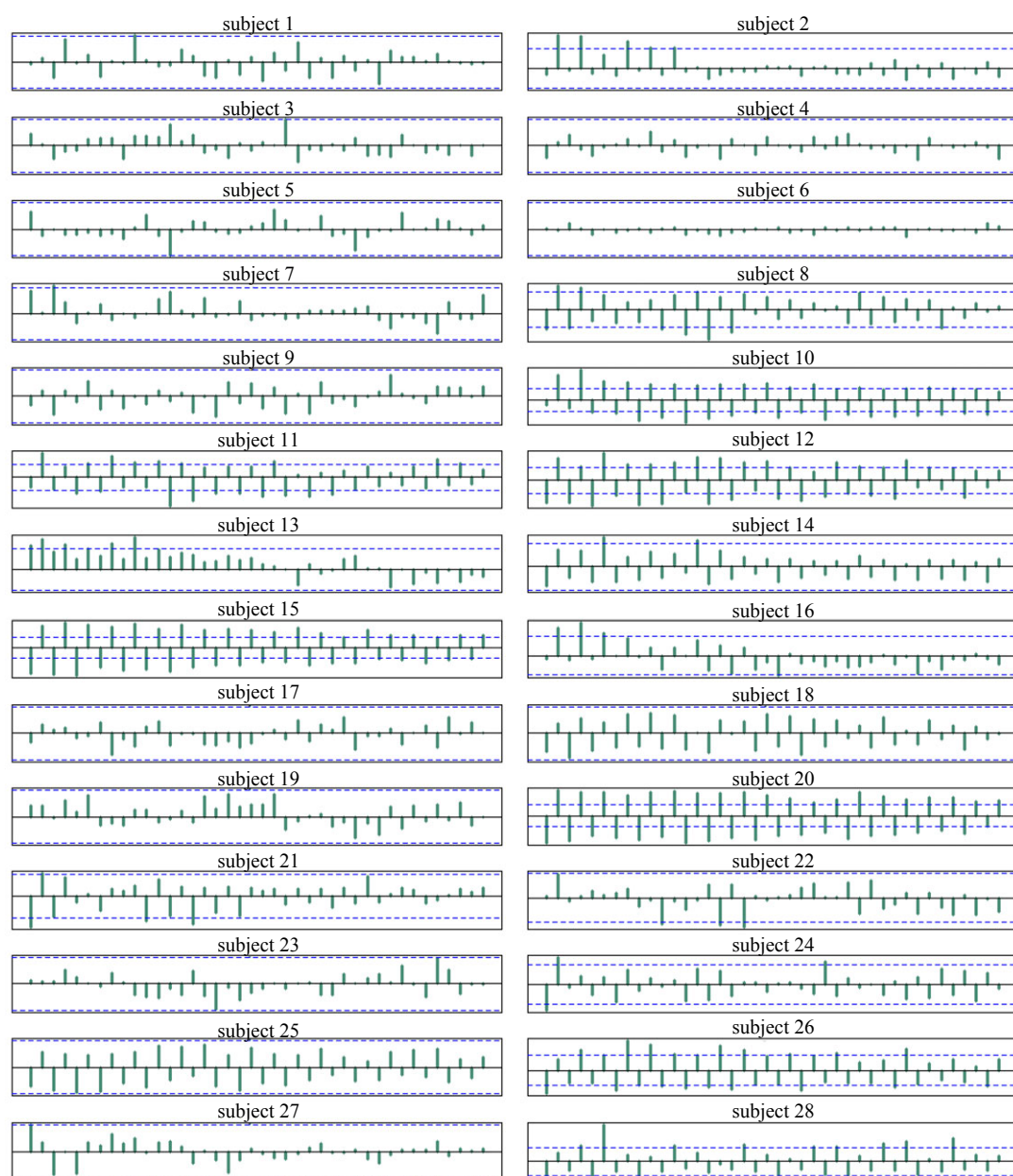

Figure 17. Autocorrelation function of $k_{\max}$ across strides for each of the 28 subjects at $v = 3.5 \text{ m s}^{-1}$. When a green vertical bar exceeding the blue dotted line in either positive or negative directions, it indicates the presence of significant autocorrelation at that lag.

k_{max} and $v = 4.5$ m s^{-1} The ACF of 4 out of 28 subjects does not seem to decay, indicating that the time series might be non-stationary.

Figure 18. Autocorrelation function of k_{max} across strides for each of the 28 subjects at $v = 4.5$ m s^{-1}.

A.4. Linear mixed effect models for each velocity, separately

The tables below show the outputs of the LMEMs built for each velocity, separately. As for the LMEMs in which both velocity and stride were included as covariates, we fitted the models using the R-package *lmer* described in [79].

All models with k_{max} (tables 9–11) show statistically significant effect of stride, while only the condition $v = 4.5$ m s^{-1} for k_{OLS} (tables 12–14) shows a statistically significant dependence on stride (see also §3.3).

Table 9. Stride has a significant fixed effect on k_{max} at $v = 2.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

k_{max} and $v = 2.5$ m s^{-1}: linear mixed effects model					
fixed effects	estimate	s.e.	d.f.	*t*-value	*p*-value
(intercept)	16218.903	330.554	28.402	49.066	**<0.0001**
stride	−4.125	1.281	2226.345	−3.221	**0.0013**

Table 10. Stride has a significant fixed effect on k_{max} at $v = 3.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

k_{max} and $v = 3.5$ m s^{-1}: linear mixed effects model					
fixed effects	estimate	s.e.	d.f.	*t*-value	*p*-value
(intercept)	16635.470	338.322	28.750	49.171	**<0.0001**
stride	−4.589	1.380	2356.552	−3.325	**0.0009**

Table 11. Stride has a significant fixed effect on k_{max} at $v = 4.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

k_{max} and $v = 4.5$ m s^{-1}: linear mixed effects model					
fixed effects	estimate	s.e.	d.f.	*t*-value	*p*-value
(intercept)	16218.903	330.554	28.402	49.066	**<0.0001**
stride	−4.125	1.281	2226.345	−3.221	**0.0013**

Table 12. Stride does not have a significant fixed effect on k_{OLS} at $v = 2.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

k_{OLS} and $v = 2.5$ m s^{-1}: linear mixed effects model					
fixed effects	estimate	s.e.	d.f.	*t*-value	*p*-value
(intercept)	22439.317	651.818	28.024	34.426	**<0.0001**
stride	−1.633	2.164	2226.256	−0.754	0.451

Table 13. Stride does not have a significant fixed effect on k_{OLS} at $v = 3.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

k_{OLS} and $v = 3.5$ m s^{-1}: linear mixed effects model					
fixed effects	estimate	s.e.	d.f.	*t*-value	*p*-value
(Intercept)	24304.192	545.335	28.384	44.567	**<0.0001**
Stride	−1.429	1.988	2356.439	−0.719	0.472

Table 14. Stride has a significant fixed effect on k_{OLS} at $v = 4.5$ m s^{-1}. (Bold indicates $p < 0.05$.)

k_{OLS} and $v = 4.5$ m s^{-1}: linear mixed effects model					
fixed effects	estimate	s.e.	d.f.	*t*-value	*p*-value
(intercept)	25089.457	588.680	28.009	42.620	**<0.0001**
stride	−3.984	1.724	2532.462	−2.311	**0.0209**

References

1. Foster KL, Higham T. 2012 How forelimb and hindlimb function changes with incline and perch diameter in the green anole, *Anolis carolinensis*. *J. Exp. Biol.* **215**, 2288–2300. (doi:10.1242/jeb.069856)

2. Hanna JB, Schmitt D. 2011 Locomotor energetics in primates: gait mechanics and their relationship to the energetics of vertical and horizontal locomotion. *Am. J. Phys. Anthopol.* **145**, 43–54. (doi:10.1002/ajpa.21465)

3. Lieberman DE, Venkadesan M, Werbel WA, Daoud AI, D'Andrea S, Davis IS, Mang'Eni RO, Pitsiladis Y. 2010 Foot strike patterns and collision forces in habitually barefoot versus shod runners. *Nature* **463**, 531–535. (doi:10.1038/nature08723)

4. Schmidt A, Fischer MS. 2011 The kinematic consequences of locomotion on sloped arboreal substrates in a generalized (*Rattus norvegicus*) and a specialized (*Sciurus vulgaris*) rodent. *J. Exp. Biol.* **214**, 2544–2559. (doi:10.1242/jeb.051086)

5. Stoessel A, Fischer MS. 2012 Comparative intralimb coordination in avian bipedal locomotion. *J. Exp. Biol.* **215**, 4055–4069. (doi:10.1242/jeb.070458)

6. Bishop M, Fiolkowski P, Conrad B, Brunt D, Horodyski M. 2006 Athletic footwear, leg stiffness, and running kinematics. *J. Athl. Train.* **41**, 387–392.

7. Foster KL, Higham T. 2014 Context-dependent changes in motor control and kinematics during locomotion: modulation and decoupling. *Proc. R. Soc. B* **281**, 201333312. (doi:10.1098/rspb.2013.3331)

8. Hancock JA, Stevens NJ, Biknevicius AR. 2007 Whole-body mechanics and kinematics of terrestrial locomotion in the elegant-crested Tinamou *Eudromia elegans*. *Ibis* **149**, 605–614. (doi:10.1111/j.1474-919X.2007.00688.x)

9. Higham TE, Jayne BC. 2004 Locomotion of lizards on inclines and perches: hindlimb kinematics of an arboreal specialist and a terrestrial generalist. *J. Exp. Biol.* **207**, 233–248. (doi:10.1242/jeb.00763)

10. Nyakatura JA, Andrada E, Grimm N, Weise H, Fischer MS. 2012 Kinematics and center of mass mechanics during terrestrial locomotion in northern lapwings (*Vanellus vanellus*, Charadriiformes). *J. Exp. Zool. Part A: Ecol. Genet. Physiol.* **317**, 580–594. (doi:10.1002/jez.1750)

11. Davis AJ, Keiter DA, Kierepka EM, Slootmaker C, Piaggio AJ, Beasley JC, Pepin KM. 2020 A comparison of cost and quality of three methods for estimating density for wild pig (*Sus scrofa*). *Sci. Rep.* **10**, 2047. (doi:10.1038/s41598-020-58937-0)

12. Cunningham TJ, Mullineaux DR, Noehren B, Shapiro R, Uhl TL. 2014 Coupling angle variability in healthy and patellofemoral pain runners. *Clin. Biomech.* **29**, 317–322. (doi:10.1016/j.clinbiomech.2013.12.008)

13. Hollander K, Liebl D, Meining S, Mattes K, Willwacher S, Zech A. 2019 Adaptation of running biomechanics to repeated barefoot running: a randomized controlled study. *Am. J. Sports Med.* **47**, 1975–1983. (doi:10.1177/0363546519849920)

14. Yaserifar M, Souza Oliveira A. 2022 Surface EMG variability while running on grass, concrete and treadmill. *J. Electromyogr. Kinesiol.* **62**, 102624. (doi:10.1016/j.jelekin.2021.102624)

15. Abourachid A. 2001 Kinematic parameters of terrestrial locomotion in cursorial (ratites), swimming (ducks), and striding birds (quail and guinea fowl). *Comp. Biochem. Physiol. Part A: Mol. Integrat. Physiol.* **131**, 113–119. (doi:10.1016/S1095-6433(01)00471-8)

16. Abourachid A, Hackert R, Herbin M, Libourel PA, Lambert F, Gioanni H, Provini P, Blazevic P, Hugel V. 2011 Bird terrestrial locomotion as revealed by 3D kinematics. *Zoology* **114**, 360–368. (doi:10.1016/j.zool.2011.07.002)

17. Bergmann PJ, Pettinelli KJ, Crockett ME, Schaper EG. 2017 It's just sand between the toes: how particle size and shape variation affect running performance and kinematics in a generalist lizard. *J. Exp. Biol.* **220**, 3706–3716. (doi:10.1242/jeb.161109)

18. Foster KL, Higham TE. 2017 Integrating gastrocnemius force–length properties, *in vivo* activation and operating lengths reveals how *Anolis* deal with ecological challenges. *J. Exp. Biol.* **220**, 796–806. (doi:10.1242/jeb.160721)

19. Irschick DJ, Jayne BC. 2000 Size matters: ontogenetic variation in the three-dimensional kinematics of steady-speed locomotion in the lizard *Dipsosaurus dorsalis*. *J. Exp. Biol.* **203**, 2133–2148. (doi:10.1242/jeb.203.14.2133)

20. Kohlsdorf T, Biewener AA. 2006 Negotiating obstacles: running kinematics of the lizard *Sceloporus malachiticus*. *J. Zool.* **270**, 359–371. (doi:10.1111/j.1469-7998.2006.00150.x)

21. Barnicoat F, Wills AP. 2016 Effect of water depth on limb kinematics of the domestic dog (*Canis lupus familiaris*) during underwater treadmill exercise. *Comp. Exercise Physiol.* **12**, 199–207. (doi:10.3920/CEP160012)

22. Buchner HHF, Savelberg HHCM, Schamhardt HC, Merkens HW, Barneveld A. 1994 Kinematics of treadmill versus overground locomotion in horses. *Vet. Q.* **16**(suppl. 2), 87–90. (doi:10.1080/01652176.1994.9694509)

23. Foster KL, Higham TE. 2010 How to build a pectoral fin: functional morphology and steady swimming kinematics of the spotted ratfish (*Hydrolagus colliei*). *Can. J. Zool.* **88**, 774–780. (doi:10.1139/Z10-043)

24. Galna B, Lord S, Rochester L. 2013 Is gait variability reliable in older adults and Parkinson's disease? Towards an optimal testing protocol. *Gait Posture* **37**, 580–585. (doi:10.1016/j.gaitpost.2012.09.025)

25. Pipkin A, Kotecki K, Hetzel S, Heiderscheit B. 2016 Reliability of a qualitative video analysis for running. *J. Orthop. Sports Phys. Therapy* **46**, 556–561. (doi:10.2519/jospt.2016.6280)

26. Sinclair J, Sant B. 2016 The effects of cross-fit footwear on the kinetics and kinematics of running. *Footwear Sci.* **9**, 41–48. (doi:10.1080/19424280.2016.1268212)

27. Spezzano LC, Jayne BC. 2016 The effects of surface diameter and incline on the hindlimb kinematics of an arboreal lizard (*Anolis sagrei*). *J. Exp. Biol.* **207**, 2115–2131. (doi:10.1242/jeb.00995)

28. Fukuchi RK, Fukuchi CA, Duarte M. 2017 A public dataset of running biomechanics and the effects of running speed on lower extremity kinematics and kinetics. *PeerJ* **5**, e3298. (doi:10.7717/peerj.3298)

29. Blickhan R. 1989 The spring-mass model for running and hopping. *J. Biomech.* **22**, 1217–1227. (doi:10.1016/0021-9290(89)90224-8)

30. McMahon TA, Cheng GC. 1990 The mechanics of running: how does stiffness couple with speed? *J. Biomech.* **23**(suppl. 1), 65–78. (doi:10.1016/0021-9290(90)90042-2)

31. Selvitella AM, Foster KL. 2022 The spring-mass model and other reductionist models of bipedal locomotion on inclines. *Integr. Comp. Biol.* **62**, 1320–1334. (doi:10.1093/icb/icac047)

32. Selvitella AM, Foster KL. 2023 Gait stability of the spring-mass model of planar locomotion on inclines. *Integr. Comp. Biol.* **62**, S99–S99.

33. Blickhan R, Full RJ. 1993 Similarity in multilegged locomotion: bouncing like a monopode. *J. Comp. Physiol. A* **173**, 509–517. (doi:10.1007/BF00197760)

34. Coleman DR, Cannavan D, Horne S, Blazevich AJ. 2012 Leg stiffness in human running: comparison of estimates derived from previously published models to direct kinematic–kinetic measures. *J. Biomech.* **45**, 1987–1991. (doi:10.1016/j.jbiomech.2012.05.010)

35. Farley CT, Glasheen J, McMahon TA. 1993 Running springs: speed and animal size. *J. Exp. Biol.* **185**, 71–86. (doi:10.1242/jeb.185.1.71)

36. Farley TF, Gonzalez O. 1996 Leg stiffness and stride frequency in human running. *J. Biomech.* **2**, 181–186. (doi:10.1016/0021-9290(95)00029-1)

37. Morin J-B, Dalleau G, Kyröl̈cainen H, Jeannin T, Belli A. 2005 A simple method for measuring stiffness during running. *J. Appl. Biomech.* **21**, 167–180. (doi:10.1123/jab.21.2.167)

38. Morin J-B, Samozino P, Zameziati K, Belli A. 2007 Effects of altered stride and contact time on leg-spring behaviour in human running. *J. Biomech.* **40**, 3341–3348. (doi:10.1016/j.jbiomech.2007.05.001)

39. Heise GD, Martin PE. 1998 Leg spring characteristics and the aerobic demand of running. *Med. Sci. Sports Exerc.* **30**, 750–754. (doi:10.1097/00005768-199805000-00017)

40. Dalleau G, Belli A, Bourdin M, Lacour JR. 1998 The spring-mass model and the energy cost of treadmill running. *Eur. J. Appl. Physiol.* **77**, 257–263. (doi:10.1007/s004210050330)

41. Müller R, Birn-Jeffery AV, Blum Y. 2016 Human and avian running on uneven ground: a model-based comparison. *J. R. Soc. Interface* **13**, 20160529. (doi:10.1098/rsif.2016.0529)

42. Blum Y, Lipfert SW, Seyfarth A. 2009 Effective leg stiffness in running. *J. Biomech.* **42**, 2400–2405. (doi:10.1016/j.jbiomech.2009.06.040)

43. Lipfert SW, Günther M, Renjewski D, Grimmer S, Seyfarth A. 2012 A model-experiment

comparison of system dynamics for human walking and running. *J. Theor. Biol.* **292**, 11–17. (doi:10.1016/j.jtbi.2011.09.021)

44. Arampatzis A, Bruggrmann G-P, Metzler V. 1999 The effect of speed on leg stiffness and joint kinetics in human running. *J. Biomech.* **32**, 1349–1353. (doi:10.1016/S0021-9290(99)00133-5)

45. Hildebrand F, Herrmann H, Fanous A. 2002 Ground reaction force and the force on center of mass. In *Int. Society of Biomechanics in Sports (ISBS 2002), Tsukuba, Japan, July 18–22, 2016*. Caceres - Extremadura, Spain.

46. Ferris DP, Louie M, Farley CT. 1998 Running in the real world: adjusting leg stiffness for different surfaces. *Proc. R. Soc. B* **265**, 989–994. (doi:10.1098/rspb.1998.0388)

47. Farley CT, Blickhan R, Saito J, Taylor CR. 1991 Hopping frequency in humans: a test of how springs set stride frequency in bouncing gaits. *J. Appl. Physiol.* **71**, 2127–2132. (doi:10.1152/jappl.1991.71.6.2127)

48. Müller R, Grimmer S, Blickhan R. 2010 Running on uneven ground: leg adjustments by muscle pre-activation control. *Hum. Mov. Sci.* **29**, 299–310. (doi:10.1016/j.humov.2010.01.003)

49. Hu Y. 2017 The role of compliance in humans and humanoid robots locomotion. PhD thesis, Heidelberg University, Heidelberg, Germany. See https://doi.org/10.11588/heidok.00022919.

50. Mombaur K. 2014 A study on optimal compliance in running. *Dynamic Walking Conference*, Zürich, Switzerland.

51. Hu Y, Mombaur K. 2016 Analysis of human leg joints compliance in different walking scenarios with an optimal control approach. *IFAC-PapersOnLine* **49**, 99–106. (doi:10.1016/j.ifacol.2016.07.992)

52. Hu Y, Mombaur K. 2017 Influence of compliance modulation on human locomotion. In *2017 IEEE Int. Conf. on Robotics and Automation (ICRA)*, Singapore, pp. 4130–4137. Piscataway, NJ: IEEE.

53. Alexander RMN. 1988 *Elastic mechanisms in animal movement*. New York, NY: Cambridge University Press.

54. Alexander RMN. 1990 Three uses for springs in legged locomotion. *Int. J. Rob. Res.* **9**, 53–61. (doi:10.1177/027836499000900205)

55. Alexander RMN. 1996 *Optima for animals*. Revised edn. Princeton, NJ: Princeton University Press.

56. Alexander RMN. 2002 *Principles of animal locomotion*. Princeton, NJ: Princeton University Press.

57. He JP, Kram R, McMahon TA. 1991 Mechanics of running under simulated low gravity. *J. Appl. Physiol.* **71**, 863–870. (doi:10.1152/jappl.1991.71.3.863)

58. Marcucci L, Reggiani C. 2020 Increase of resting muscle stiffness, a less considered component of age-related skeletal muscle impairment. *Eur. J. Transl. Myol.* **30**, 223–233. (doi:10.4081/ejtm.2020.8982)

59. Quatman CE, Wiseman J, Phieffer L. 2018 Academic geriatric orthopaedics: a new paradigm for inpatient care. *Curr. Geriatr. Rep.* **7**, 272–277. (doi:10.1007/s13670-018-0266-9)

60. Xu J, Fu SN, Hug F. 2021 Age-related increase in muscle stiffness is muscle length dependent and associated with muscle force in senior females. *BMC Musculoskelet. Disord.* **22**, 1–7.

61. Lark SD, Buckley JG, Bennet S, Sargeant AJ. 2003 Joint torques and dynamic joint stiffness in elderly and young men during stepping down. *Clin. Biomech.* **18**, 848–855. (doi:10.1016/S0268-0033(03)00150-5)

62. Carruthers AS, Farley CT. 1998 Leg stiffness in running humans: effects of body size. *In Proc. of The Third North American Congress on Biomechanics (NACOB '98), Waterloo, Ontario, Canada*, pp. 307–308.

63. Farley CT, Korff WL. 1999 Musculoskeletal basis for the scaling of leg stiffness with body mass in humans. *In Proc. of the 23rd Annual Meeting of the American Society of Biomechanics, Pittsburgh, PA. October 21–23 1999*.

64. Fox MC, Hsiao-Wecksler ET, Polk JD. 2022 Effects of body mass on leg and vertical stiffness in running humans. *bioRxiv*: https://www.biorxiv.org/content/10.1101/2022.05.04.490639v1.full.pdf.

65. Herr HM. 1998 A model of mammalian quadrupedal running. PhD thesis, publication no. 9822898. Harvard University, Harvard, Cambridge, MA, USA. See https://search.proquest.com/dissertations/docview/304436279/abstract/4201D4C3C947449FPQ/1.

66. Ernst M, Götze M, Müller R, Blickhan R. 2014 Vertical adaptation of the center of mass in human running on uneven ground. *Hum. Mov. Sci.* **38**, 293–304. (doi:10.1016/j.humov.2014.05.012)

67. Casella G, Berger RL. 2002 *Statistical inference*, 2nd edn. Belmont, MA: Cengage Learning.

68. Yule GU. 1921 On the time-correlation problem, with especial reference to the variate-difference correlation method. *J. R. Stat. Soc.* **84**, 497–537. (doi:10.2307/2341101)

69. Yule GU. 1926 Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *J. R. Stat. Soc.* **89**, 1–63. (doi:10.2307/2341482)

70. Anderson TW, Walker AM. 1964 On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *Ann. Math. Stat.* **35**, 1296–1303. (doi:10.1214/aoms/1177703285)

71. Hannan EJ, Hyde CC. 1972 On limit theorems for quadratic functions of discrete time series. *Ann. Math. Stat.* **43**, 2058–2066. (doi:10.1214/aoms/1177690885)

72. Bartlett MS. 1946 On the theoretical specification and sampling properties of autocorrelated time-series. *Suppl. J. R. Stat. Soc.* **9**, 27–41. (doi:10.2307/2983611)

73. Box GEP, Jenkins GM. 1970 *Time series analysis: forecasting and control*. San Francisco, CA: Holden-Day.

74. Box GEP, Pierce DA. 1970 Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.* **65**, 1509–1526. (doi:10.1080/01621459.1970.10481180)

75. Ljung GM, Box GEP. 1978 On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303. (doi:10.1093/biomet/65.2.297)

76. Laird NM, Ware JH. 1982 Random-effects models for longitudinal data. *Biometrics* **38**, 963–974. (doi:10.2307/2529876)

77. Bartlett MS. 1937 Properties of sufficiency and statistical tests. *Proc. R. Soc. A* **160**, 268–282. (doi:10.1098/rspa.1937.0109)

78. Patterson HD, Thompson R. 1971 Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554. (doi:10.1093/biomet/58.3.545)

79. Bates D, Mächler M, Bolker B, Walker S. 2015 Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48. (doi:10.18637/jss.v067.i01)

80. Holmes P, Full RJ, Koditschek DE, Guckenheimer J. 2006 The dynamics of legged locomotion: models, analyses, and challenges. *SIAM Rev.* **48**, 207–304. (doi:10.1137/S0036144504445133)

81. Ghigliazza RM, Holmes P. 2005 Towards a neuromechanical model for insect locomotion: hybrid dynamical systems. *Regul. Chaotic Dyn.* **10**, 193–225. (doi:10.1070/RD2005v010n02ABEH000311)

82. Fan J, Yao Q. 2003 *Nonlinear time series - nonparametric and parametric methods*. New York, NY: Springer.

83. Nyamundanda G, Poudel P, Patil Y, Sadanandam A. 2017 A novel statistical method to diagnose, quantify and correct batch effects in genomic studies. *Sci. Rep.* **7**, 10849. (doi:10.1038/s41598-017-11110-6)

84. Panda R, Kalmady SV, Greiner R. 2022 Multi-source domain adaptation techniques for mitigating batch effects: a comparative study. *Front. Neuroinformat.* **16**, 805117. (doi:10.3389/fninf.2022.805117)

85. Foster KL, Selvitella AM. 2020 Learning the locomotion behaviour of lizards transfers across environments. In *ICML 2020 Workshop on Computational Biology* (spotlight presentation).

86. Foster KL, Selvitella AM. 2022 Transfer of anolis locomotor behavior across environments and species. *Integr. Comp. Biol.* **62**, 774–790. (doi:10.1093/icb/icac015)

87. Foster KL, Selvitella AM. 2023 Anolis ecomorph biomechanics across arboreal environments: what can machine learning tell us about behavioral plasticity in lizards? *Integr. Comp. Biol.* **62**, S99–S99.

88. Button K, Ioannidis J, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafó MR. 2013 Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376. (doi:10.1038/nrn3475)

89. Ioannidis JPA. 2005 Why most published research findings are false. *PLoS Med.* **2**, e124. (doi:10.1371/journal.pmed.0020124)

90. Basu D. 1955 On statistics independent of a complete sufficient statistic. *Sankhy* **15**, 377–380.

91. Grubbs FE. 1950 Sample criteria for testing outlying observations. *Ann. Math. Stat.* **21**, 27–58. (doi:10.1214/aoms/1177729885)

92. Frees E. 2010 *Regression modeling with actuarial and financial applications*, 1st edn. New York, NY: Cambridge University Press.

93. Selvitella AM, Foster KL. 2023 On the variability and dependence of human leg stiffness across strides during running and some consequences for the analysis of locomotion data. Figshare. (doi:10.6084/m9.figshare.c.6777969)