On Gait Consistency Quantification Through ARX Residual Modeling and Kernel Two-Sample Testing

A. Stihi, T. J. Rogers, C. Mazzà, and E. J. Cross

Abstract*—*Objective: **The quantification of the way an individual walks is key to the understanding of diseases affecting the neuromuscular system. More specifically, to improve diagnostics and treatment plans, there is a continuous interest in quantifying gait consistency, allowing clinicians to distinguish natural variability of the gait patterns from disease progression or treatment effects. To this end, the current article presents a novel objective method for assessing the consistency of an individual's gait, consisting of two major components.** ***Methods:*** **Firstly, inertial sensor accelerometer data from both shanks and the lower back is used to fit an AutoRegressive with eXogenous input model. The model residuals are then used as a key feature for gait consistency monitoring. Secondly, the non-parametric maximum mean discrepancy hypothesis test is introduced to measure differences in the distributions of the residuals as a measure of gait consistency. As a paradigmatic case, gait consistency was evaluated both in a single walking test and between tests at different time points in healthy individuals and those affected by multiple sclerosis (MS).** ***Results:*** **It was found that MS patients experienced difficulties maintaining a consistent gait, even when the retest was performed one-hour apart and all external factors were controlled. When the retest was performed one-week apart, both healthy and MS individuals displayed inconsistent gait patterns.** ***Conclusion:*** **Gait consistency has been successfully quantified for both healthy and MS individuals.** ***Significance:*** **This newly proposed approach revealed the detrimental effects of varying assessment conditions on gait pattern consistency, indicating potential masking effects at follow-up assessments.**

***Index Terms*—Gait consistency, Auto-Regressive, maximum mean discrepancy (MMD), multiple sclerosis (MS), wearable sensors.**

Manuscript received 27 February 2023; revised 20 June 2023 and 10 August 2023; accepted 7 September 2023. Date of publication 18 September 2023; date of current version 26 February 2024. The work of T. J. Rogers was supported by EP/W002140/1. The work of E. J. Cross and T. J. Rogers was supported by EP/W005816/1. This work was supported in part by the National Institute for Health Research (NIHR) through the Sheffield Biomedical Research Centre (BRC), under Grant IS-BRC-1215-20017, and in part by the Innovative Medicines Initiative 2 Joint Undertaking (JU) through the Mobilise-D Project under Grant 820820. This JU was supported by the European Union's Horizon 2020 Research and Innovation Program and the European Federation of Pharmaceutical Industries and Associations (EFPIA). *(Corresponding author: A. Stihi.)*

A. Stihi is with the Department of Mechanical Engineering, Insigneo Institute for in silico Medicine and Dynamics Research Group, University of Sheffield, S10 2TN Sheffield, U.K. (e-mail: astihi1@sheffield.ac.uk).

T. J. Rogers and E. J. Cross are with the Department of Mechanical Engineering and the Dynamics Research Group, University of Sheffield, U.K.

C. Mazzà is with the Department of Mechanical Engineering and Insigneo Institute for in silico Medicine, University of Sheffield, U.K.

Digital Object Identifier 10.1109/TBME.2023.3316474

I. Introduction

GAIT is a multifaceted dynamic process, involving the coordination of the lower limbs as well as the upper body [1]. Scientific literature supports the notion that variations from the "normal walking patterns" may indicate certain pathological conditions [2]. While methods for identifying gait impairment have been well-established in research, less attention has been given to the examination of the longitudinal progression of gait patterns over time [3].

From an energetic perspective [4], the ideal gait pattern would be identical across all steps. However, it is well established that in reality, gait only displays approximate periodicity [5], and it can also be altered over time [6]. Therefore, the authors propose that the longitudinal quantification of gait may be significantly impacted by the natural fluctuations in testing conditions during clinical follow-up assessments. Some causes may include marginal discrepancies in sensor attachment locations on body segments, time of assessment in the presence of medications, etc. These might mask subtle degradation or rehabilitation in pathological populations. As a result of the variability introduced by these differences, accurately quantifying the longitudinal gait changes, while removing the influence of the confounding factors, can be a very difficult task. As such, to improve diagnostics and treatment outcomes, there is a need for an objective gait consistency measure, which can quantify one individual's ability to consistently repeat the same walking pattern, regardless of the environment or task. This consistency measure can be seen as a marker of good motor control and balance, and the lack of it may indicate neurological or musculoskeletal problems.

Recent advancements in wearable technology, specifically inertial measurement units (IMUs), have increased their use in clinical gait analysis [7]. Due to their flexibility, IMUs have the potential of being practical solutions for quantifying gait consistency, allowing clinicians to examine the dynamic link between the lower limbs and the upper body movements during walking tests.

In view of the observed variability in the gait pattern, owing to both environmental and pathological influences, the current work introduces a novel measure for the consistency of an individual's gait. This new tool provides an objective method for assessing the variability in the gait both within a single walking test and between tests. This is achieved by means of two novel components. Firstly, the residuals of a dynamic AutoRegressive with eXogenous input (ARX) model [8] between both shanks and the lower back are used as a sensitive feature. Secondly, the maximum mean discrepancy (MMD) [9] is introduced to measure the differences in the distribution of the residuals. This approach leads to a sensitive and informative method for evaluating the consistency of the gait patterns.

A. Related Work

In the field of gait analysis, the term "gait consistency" has been utilized with various interpretations. Some literature defines it as the accuracy of the measurement tools employed to obtain gait data [7], [10], [11], [12], evaluated by comparing specific gait features between different assessments. Conversely, other authors refer to consistency as the regularity of the gait patterns that recur in each gait cycle [5], [6]. It is important to note that consistency pertains to the similarity of gait patterns over a period of time [13], and therefore having a clear understanding of the concept of gait consistency is essential in order to differentiate between natural variability[1] disease progression or treatment effects.

Maintaining proper coordination between the lower limbs and the upper body is essential for consistent gait patterns [1]. The coordination of the trunk and upper body is a requirement in order to maintain balance and stability during gait [1], [14], [15]. Similarly, good coordination of the hip, knee and ankle joints is necessary for proper weight bearing and forward propulsion during walking [16], [17]. In contrast, the lack of coordination can manifest through inconsistent gait patterns, which can also be seen as a result of a specific pathological condition [2], or associated with a high risk of falls [18], [19]. Recent developments in wearable technology have facilitated the use of inertial measurement units (IMUs) to effectively monitor the dynamic relationship between the lower limbs and upper body [16], [20]. In the relevant literature, numerous studies have focused on extracting spatio-temporal metrics from the gait signals [7], [10], [14], [20], [21], [22], [23], [24]. Although, this is a viable approach, it has the major disadvantage of potentially discarding a significant amount of information which may be beneficial for quantifying the consistency of the gait patterns [14]. Several other models have been proposed in the literature to describe the basic acceleration patterns of the pelvis using physics-informed models, including the inverted pendulum model proposed by [25] or the dynamic walking perspective proposed by [4]. Yet, these types of modelling approaches are a simplification of the actual gait dynamics. Despite the fact that they can be used to accurately estimate spatio-temporal parameters in healthy populations [23], they do not work as well for more complex, pathological gait patterns. As a result, the gait periodicity and symmetry are inaccurately estimated [26].

This article deals with the use of a data-based modelling approach, as it can provide a potentially effective alternative for monitoring the progression of the disease. The model relies on acceleration measurements taken at the beginning of a walking test. If properly established, the authors hypothesise that such a model would be able to reveal changes in the gait patterns either along the walking test, or at a later stage, longitudinally, in the event of changes in patient's health status. This approach has been inspired by the structural health monitoring (SHM) field [27], and it leverages an ARX model, which is a linear representation of a dynamic system in discrete time. Autoregressive models have been successfully deployed for structural damage identification using accelerometer data in [28], [29], [30], [31]. In SHM, damage detection algorithms can be constructed by examining the residual error (i.e., the difference between the measured data and the model predictions) as a damage-sensitive feature. Inspired by the concept of monitoring the progress of the residual error as an indicator of change, the authors believe that the same approach could work in the field of gait analysis. The goal is to reveal gait anomalies present in the distribution of the residuals, which could be caused by gait disabilities due to the presence of some disease. Consequently, establishing whether a significant change occurred in the residual patterns gives raise to the requirement of statistical hypothesis testing. Fortunately, a suitable objective hypothesis testing method is represented by the MMD two-sample hypothesis test, which will be introduced later in Section IV. Thus, this will contribute to the understanding of gait consistency in both healthy and pathological subjects, as well as quantifying the influence of varying environmental testing conditions.

The article is structured as follows. Firstly, the overview of the approach is presented with in Section II. Then, an example dataset used to demonstrate the approach is introduced in Section III. In Section IV, the relevant background is presented for the residual modelling procedure, followed by the details of the implementation of the MMD-based hypothesis test. Then, Section V presents the application of the novel methodology on the dataset presented in Section III. The implications of the methodology are further discussed in Section VI. Finally, the article concludes in Section VII.

II. Overview of the Novel Approach for Assessing Gait Consistency

As previously described, the approach for quantifying gait consistency introduced in this article consists of two main components. The first component entails modelling the dynamic relationship between the upper body and lower limb acceleration patterns, from which the residual patterns are obtained. The second component of the process consists in the introduction of a novel hypothesis test in the field of gait analysis, used for the quantitative estimation of gait consistency. In this section, a high-level overview of the approach will be presented to the reader, while the detailed modelling procedures are postponed until Section IV.

[1]For clarification, even though the term "natural variability" is usually employed to denote the intrinsic variability necessary to maintain balance and adapt to environmental changes, here it is used solely to denote inherent fluctuations in the walking patterns recorded over a period of time.

Following the successful data-based inference approaches in SHM, the authors propose to capture the dynamic relationship between the upper body and lower limb movements using a similar modelling procedure based on ARX-type models [8]. In such a model, the output of the model is a linear function of previous or lagged instances of the output and lagged instances of the inputs. Concerning gait, the trunk acceleration displays a pseudo-periodic pattern which approximately repeats every gait cycle [5]. As a result, the acceleration measurements will be autocorrelated, preventing the use of any statistical methodology that ignores correlation [32]. If the correlation is not removed from the original signals, the anomaly detector may trigger many false alarms and might fail to recognise any anomalies present due to the presence of the disease, or indeed due to the inability to walk consistently during walking tests. However, monitoring the residuals obtained by fitting an ARX-type model to the observed data can overcome this problem. In this case, the output of the model is taken as the 3-axial acceleration norm measured at the lower back, while the two inputs are taken as the acceleration norm measured at the shanks. If the ARX-type model is a reasonably accurate representation of the system being modelled, then, the model residuals should exhibit little to no correlation and should appear to be white noise, lacking any discernible systematic pattern.

The philosophy introduced by this modelling strategy is straightforward: if one identifies a good time-series model for a particular individual, and, if the respective individual has a stable and controlled gait, i.e. it remains consistent, then the model will make good predictions. As a result, the residuals should have a low variance and be centred at zero. If the system response is changing, for example due to balance and coordination deficits, then the previously identified model will not be able to make good predictions and the variance of the residual sequence will increase (relative to the stable condition). When inconsistencies are identifiable within the residual patterns, something has changed. Various sources can be responsible for these changes, such as changes in health condition, differences in testing conditions, etc.

Having established the modelling strategy, it is then important to decide the objective strategy for determining the similarity of the residuals. For this reason, an objective statistical measure is a necessity, along with the corresponding hypothesis test. The test used in this work is revolved around the kernel-based [33] computation of the Maximum Mean Discrepancy [9]. Here, a kernel-based implementation only requires the user to specify a similarity function, formulated as an inner product in a feature space, which is infinite dimensional[2] [33]. The full motivation for adopting the MMD as the preferred statistical measure is postponed until Section IV.

Once the statistical metric has been established, the only final requirement is to transform the associated hypothesis test into an objective measure which can then be used to quantitatively report the consistency of gait. Among the other reasons specified in Section IV, one can simply cross-compare smaller segments of the residual patterns and assess their similarity, by setting up a hypothesis test. The latter makes use of the MMD metric and evaluates the dissimilarity of the distributions between two groups by measuring the discrepancy between samples drawn from each distribution. Then, MMD test statistic is computed as the difference between Hilbert space embeddings [34] of the two sets of samples being compared. If the difference is large, then it is likely that the two residual distributions are different and there is an inconsistency in gait. The total number of comparisons will be equal to the square of total number of smaller residual segments (for the graphical illustration refer to Fig. 3). Implementing the hypothesis test across all combinations of residual segments can then lead to creation of *accept-reject maps*, as depicted in Fig. 4. The results presented in the flowchart diagram (Fig. 1) are shown purely for their qualitative aspects and the full explanation and interpretation will be provided at a later stage in this article. Finally, the consistency of gait can then be computed as the percentage of the number of times the hypothesis test regarded the distribution of the residual segments as being different relative to the total number of comparisons.

III. A Case-Study to Demonstrate the Proposed Approach

A. Participants

The dataset used in this work consists of inertial measurement unit (IMU) acceleration recordings from 2 groups of subjects. The first group (group A) consisted of 14 healthy controls (HCs, with no history of musculoskeletal or neurological disorders which might affect their balance or mobility) and 26 individuals with multiple sclerosis (MS). The latter is a neurogenerative disease characterised by the inflammatory-mediated demyelination of the axons in the central nervous system [35], primarily manifesting through gait balance and coordination deficits [20], [35]. The first group of subjects completed the baseline test and the retest on the same day, one hour apart. The sensors were not repositioned between the two tests. An additional group (group B) consisting of 23 HCs and 24 individuals with MS performed the retest one week apart from the baseline test (see Table I). This was done to introduce natural variability in changing assessment conditions and its effect on gait consistency within a period in which the disease status would not change. The severity of the disease for the patients with MS was assessed through the Expanded Disability Status Scale (EDSS) [36]. The subjects with relapse-remitting MS were only included in the study if no relapse occurred for 30 days prior to the baseline test and had stable treatments for the past three months. The study was approved by the NRES Committee Yorkshire & The Humber-Bradford Leeds (Ref: 15/YH/0300) and by the North of Scotland Research Ethics Committee (Ref: 17/NS/0020). All subjects provided written informed consent before entering the study.

[2]Although there are a large number of kernels available, in practice, certain options are frequently used due to their general applicability. For instance the Gaussian/RBF kernel is a widely adopted choice, and it is also the kernel function employed in this work.

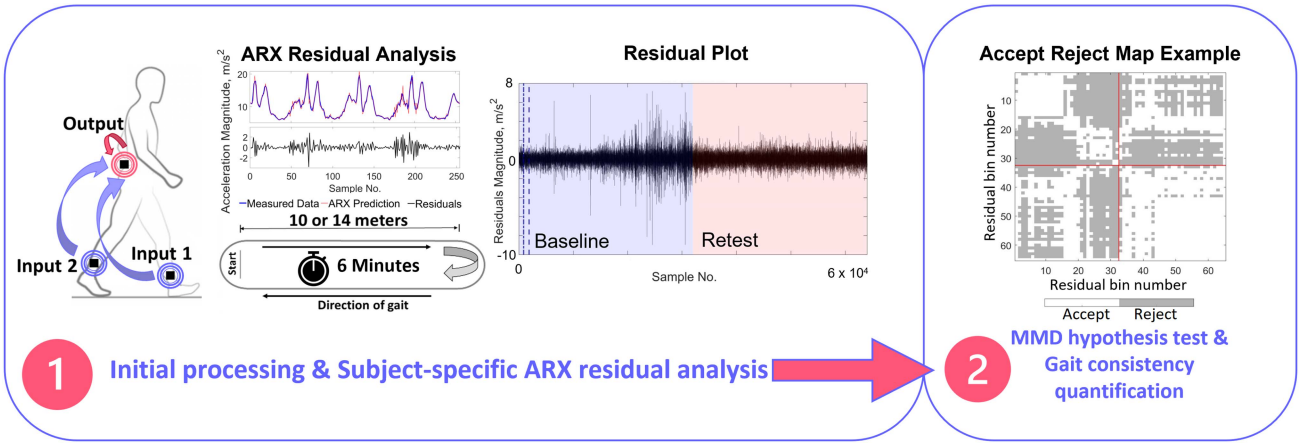

Fig. 1. Flowchart of the modelling approach.

TABLE I
DEMOGRAPHICS TABLE

		Age	Gender	MS Subtypes			Walking assistive devices		
		Mean (SD)	*N male*	*PP*	*RR*	*SP*	*None*	*Unilateral*	*Bilateral*
Group A	**HC** (n=14)	27.4 (3.7)	8	-	-	-	14	0	0
	MS (n=26) $\overline{EDSS} = 3.9$	44 (13.1)	5	2	23	1	19	5	2
Group B	**HC** (n=23)	49.4 (8.0)	7	-	-	-	23	0	0
	MS (n=21) $\overline{EDSS} = 5.1$	56.5 (10.4)	6	0	0	21	10	8	3

PP = primary progressive, RR = relapse remitting, SP = secondary progressive

B. Gait Assessment and Initial Processing

Gait data was collected using three tri-axial IMUs, (OPAL, APDM Inc, Portland, OR, USA, sampling frequency, 128 Hz, accelerometer range ±6 g), attached to the body through elastic straps, on the anterior aspect of both lower shanks and on the lower back (L4 – L5). The sensors were configured for synchronised recording using the manufacturer's provided access point. The sensing axes of the sensors were approximately aligned to the anatomical planes. Group A performed the baseline test by walking along a 14-m corridor, while the group B walked along a 10-m corridor, going back and forth for 6 minutes. The schematic of the testing procedure is illustrated in Fig. 1. All participants were instructed to walk at their self-selected pace. Resting was allowed, if needed. Additionally, walking aids were permitted, only if used daily.

Following the procedure detailed in [20], all turns and resting breaks were automatically removed, and only straight-line walking bouts of steady-state walking were included in the following analysis. To mitigate the end effects before and after the turns, the first and last strides in a walking bout were also excluded from the analysis. The gait events were identified from the shank angular velocity signals, as described in [20]. For the sake of brevity, the details of the gait events algorithms used are not replicated here. To avoid the effects of possible undesired minor movements of the sensors between sessions, it was decided to only work with the raw 3-axial acceleration norm signals. For full details of the data collection and pre-processing, the readers are referred to [20].

IV. MEASURING GAIT CONSISTENCY

A. Part 1 - ARX Time Series Residual Modelling

Time series analysis represents a statistical framework concerning the extraction of significant statistics or characteristics of the sequence of observations for various purposes, ranging from model identification to forecasting of future values from current and past values [37]. As discussed in the previous sections, the entire philosophy introduced by the first part of this work consists of monitoring the residual sequences once an ARX model is fit to the gait data. It should also be noted that the following section is a specific introduction only to the modelling strategy used in this work. For a more comprehensive overview regarding time-series modelling approaches the reader is referred to [8]. Here, the ARX model can be written in the following form, where the output y at time t, is given by:

$$y(t) = \sum_{i=1}^{n_a} a_i y(t-1) + \sum_{j=1}^{2} \sum_{k=1}^{n_b(j)+1} b_j(k-1)u_j \times (t-k+1) + e(t) \quad (1)$$

where n_a is the number of lags for the output (in this case the lower back 3-axial acceleration norm), $n_b(j)$ is the number lags for the corresponding input (the left or right shank 3-axial acceleration norm), a_i is the i-th output coefficient, u_j is the j-th system input and its corresponding coefficient is b_j. Finally, the noise is represented by $e(t)$. Note that the inputs also contain the static regression, as the lower limb and upper body movements

occur simultaneously. Additionally, (1) is only valid for one-step ahead predictions. It is also noteworthy that the modelling procedure proposed here is only applicable if the sampling rate (which is defined as the inverse of the time interval between two consecutive measurements, and determines the temporal resolution of the data) is maintained constant between baseline and follow-up assessments. In the case of this work, the sampling rate has been kept constant at 128 Hz across all measurements.

The parameters of the ARX model can be fit to minimise the one step ahead prediction error using ordinary least squares. Remembering that the dataset used in this work consists of individuals performing a walking test, going back and forth along a straight corridor, the coefficients were computed using the gait acceleration signals measured during the first straight-line walking bout for all combinations of model orders, where n_a and n_b were varied from 1 to 15. The model order can then be selected using the Bayesian Information Criterion (BIC) [38] on a validation set. In the case of this work, the validation set consisted in gait data measured during the second straight line walking bout. The preferred model is then selected as the one corresponding to the minimum BIC, which is defined as:

$$BIC = -2\ln(\hat{L}) + p\ln(N) \tag{2}$$

where $\hat{L}$ is the maximum value of the likelihood estimate of the model tested, given the data, p denotes the number of parameters used by the model, and N is the number of observations. The BIC allows for the comparison of different model structures. It was deemed an appropriate model selection criterion, as it introduces a penalty term for the number of parameters used in the model, reducing the chances of overfitting [39].

The next step in the ARX processing workflow consisted of computing the model residuals vector, $\mathbf{R}$, calculated as the difference between the measured data and the model prediction, as indicated by (3). Here, $\mathbf{Y}$ is the output vector, X is the input matrix and $\hat{\theta}$ is the ARX coefficients vector.

$$\mathbf{R} = \mathbf{Y} - X\hat{\theta} \tag{3}$$

Once an appropriate model is selected, the residuals are computed across all the remaining straight-line walking bouts of the baseline test, as well as during all the straight-line walking bouts of the retest. This assessment was done in order to capture any variations in the system dynamics during the retest and to allow the quantification of consistency of the gait patterns. Given the test-retest availability of data, the main benefit of this approach is that once the orders and coefficients of the ARX models have been established at the baseline, they do not have to be determined again at a later stage. As will be discussed in the following sections of the article, this condition will prove to be extremely useful at highlighting the influence of the confounding factors. Therefore, extracting the model residuals represents the foundation of the workflow presented in this article. The second part of the workflow consists of monitoring the model residuals, assuming that the residual pattern across a walking test remains similar, given that the person performing the test has a stable and controlled gait, i.e. it is consistent. The determination of whether two sequences of residuals are consistent across different time points involves statistical hypothesis testing, which provides a means of quantifying the degree of gait consistency.

B. Residual Pattern Comparison – the Maximum Mean Discrepancy as the Preferred Statistical Metric

Assumptions regarding the form of the residual distributions should not be made without scrutiny. Consequently, it becomes a necessity to explore flexible approaches to accurately quantify the discrepancies observed between repeated measurements and provide an objective comparison between the test and retest residual patterns. To this end, a statistical metric is required, along with a corresponding hypothesis test, which should:

1) quantify the differences between residual patterns using an objective approach, while providing the clinicians with consistent and interpretable results.
2) account for the complete form of the distribution, and not just a small number of statistical moments.
3) provide non-parametric estimations with convergence guarantees for the density estimations, so that it can be applied to any given distributions.

While the first point on the requirements list is obvious and is targeting operator bias, the following two need further introduction. Due to the natural variability in the gait patterns in both healthy and pathological populations, no two residual distributions must be assumed to be the same. Therefore, this inevitably raises some problems since no a-priori knowledge about the form of the residual distributions should be assumed. The kernel trick offers a solution to this problem by allowing for the assessment of an infinite number of statistical moments through the use of inner products in a feature space [33]. This approach extends the comparison of distributions, rather than specifying which features to focus on in advance.

The Maximum Mean Discrepancy (MMD) [9] is a metric which fulfils all the requirements specified above and is defined as the maximum difference between the mean kernel embeddings of features [40]. Specifically, the unbiased MMD can be formally estimated as follows:

$$\begin{aligned} MMD^2(\mathbb{X},\mathbb{Y}) = &\frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j\neq i}^{m} k(x_i,x_j) \\ &+ \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} k(y_i,y_j) \\ &- \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(x_i,y_j) \end{aligned} \tag{4}$$

where $\mathbb{X}$ and $\mathbb{Y}$ are the two residual distributions to be compared, x_i and y_i are samples drawn from these distributions, and m and n are the corresponding sample sizes of $\mathbb{X}$ and $\mathbb{Y}$ respectively. To clarify, $\mathbb{X}$, and $\mathbb{Y}$ are stated as probability measures, but generally, will be utilised in the form of a probability density function (PDF).

Although, there are many kernel types that can be selected for the computation of the MMD, one of the most popular choices is represented by the radial basis function (RBF) kernel [9], [41],

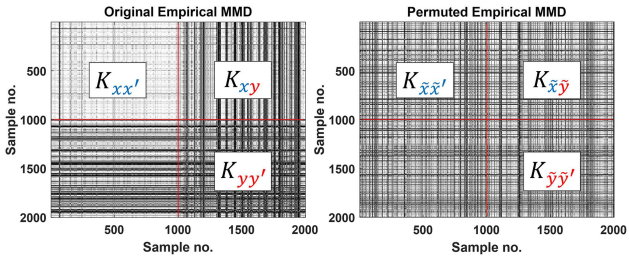

Fig. 2. Visualisation of the aggregate sample matrix KZ and the permuted matrix, KZ_{perm}. The first 1000 samples correspond to a region of 1000 datapoints within a residual signal, while the last 1000 correspond to a different sequence of 1000 datapoints within a different residual signal. Here, x and x' are independent variables with distribution $\mathbb{X}$, y and y' are also independent variables with distributions $\mathbb{Y}$, where x' and y' are independent copies of x and y within the same distributions. $\tilde{x}$ and $\tilde{y}$ are independent variables drawn from the permuted distributions $\tilde{\mathbb{X}}$ and $\tilde{\mathbb{Y}}$. The red lines are used just for delimitation purposes.

defined as:

$$k(x,y) = \exp\left(-\frac{||x-y||^2}{2\sigma^2}\right) \tag{5}$$

where σ is the parameter controlling the of bandwidth of the kernel. While Gretton et al. [42] suggest setting up the kernel bandwidth using the median heuristic of the aggregate sample (i.e., concatenating the two datasets into a single sample), there is no theoretical understanding of when this is a good choice, and in some cases, it might not be the optimal solution [43], [44]. Therefore, an optimization procedure would be better suited for this application, which is to be discussed in the upcoming paragraphs.

C. MMD Hypothesis Test

To quantify the consistency of gait, one could pose the question of whether the two residual distributions to be compared are similar, and set up a hypothesis test. Given a set of independent observations drawn from two distributions $\mathbb{X}$ and $\mathbb{Y}$, the hypothesis test is used to differentiate between the null hypothesis and the alternative hypothesis via comparison of the test statistic with a particular threshold. Here, the null hypothesis was set up as H_0: $\mathbb{X} = \mathbb{Y}$, whereas the alternative hypothesis is H_1: $\mathbb{X} \neq \mathbb{Y}$. The procedure detailed in Algorithm 1, involving the MMD-based hypothesis test is presented here for completeness. Setting up the hypothesis test starts by computing the kernel embedding matrices $K_{xx'}$, $K_{yy'}$ and K_{xy}, followed by the computation of the MMD test statistic, according to (4). Then, the aggregate sample matrix, KZ, is formed, as detailed in Line 3, Algorithm 1. This pre-computation avoids the quadratic-time computational cost in the forthcoming permutation loop. In order to create artificially symmetric distributions, a bootstrapping procedure is employed to establish the objective threshold (see Fig. 2) [9]. This procedure is performed a specified number of times, indicated by the variable "*no permutations*". During each iteration, the MMD is recalculated for the permuted distributions. Finally, the distances are then sorted in ascending order of magnitude and the threshold is selected as the distance corresponding to the desired confidence level. If the threshold is exceeded by the test statistic, then the test rejects the null hypothesis, as there is not enough evidence to believe that samples x and y were drawn from the same distribution. Otherwise, the null hypothesis that the two distributions are the same is accepted.

Algorithm 1: MMD Hypothesis Test.

1: Compute $K_{xx'} = k(x,x')$, $K_{yy'} = k(y,y')$ and $K_{xy} = k(x,y)$
2: Compute the MMD test statistic as: $testStat = \mathbb{E}[K_{xx'}] + \mathbb{E}[K_{yy'}] - 2\mathbb{E}[K_{xy}]$
3: Store $KZ = \begin{pmatrix} K_{xx'} & K_{xy} \\ K_{yx} & K_{yy'} \end{pmatrix}$
4: **for** $i = 1 : no\ permutations$ **do**
5: Permute elemets of KZ and construct: $KZ_{perm} = \begin{pmatrix} K_{\tilde{x}\tilde{x}'} & K_{\tilde{x}\tilde{y}} \\ K_{\tilde{y}\tilde{x}} & K_{\tilde{y}\tilde{y}'} \end{pmatrix}$
6: Compute the permuted MMD as: $MMD_{perm} = \mathbb{E}[K_{\tilde{x}\tilde{x}'}] + \mathbb{E}[K_{\tilde{y}\tilde{y}'}] - 2\mathbb{E}[K_{\tilde{x}\tilde{y}}]$
7: Store MMD_{perm} in MMD_{perm_array}
8: **end for**
9: Sort MMD_{perm_array}
10: Compute $threshold$ as the distance corresponding to the desired confidence level
11: **if** $testStat > threshold$ **then**
12: Reject null hypothesis: H_1: $\mathbb{X} \neq \mathbb{Y}$
13: **else**
14: Accept null hypothesis: H_0: $\mathbb{X} = \mathbb{Y}$
15: **end if**

D. MMD Kernel Bandwidth Optimization

From (5) it is noted that the bandwidth hyperparameter (σ, controlling the width of the kernel) needs to be objectively established. As mentioned previously, setting up σ as the median pairwise distance among the aggregate sample is not desirable. Gretton et al. [9] showed that this method is not suitable for large datasets, since it can lead to significant type-II errors (i.e. when the two distributions are regarded as being the same by the hypothesis test, despite being different in reality). In a different study, Gretton et al. [44] proposed an optimization procedure for large sample sets, by selecting linear combinations of kernels which minimize type-II errors and therefore improve the robustness to false negatives of the MMD when used as a test-statistic for a two-sample hypothesis testing. Moreover, the RBF kernel embedding allows for an increased resolution for characterising any given distributions. This effectively translates into the embedding of an infinite dimensional vector of statistical moments, resulting in an asymptotic guarantee that the hypothesis test will capture any distribution and become overly sensitive to the infinitely small differences, which is not appropriate.

Noting the above observations and recognising that, in practice, the MMD is rarely applied to large datasets consisting of more than a couple thousand data points, the residual signals were divided into smaller segments consisting of 1000 data

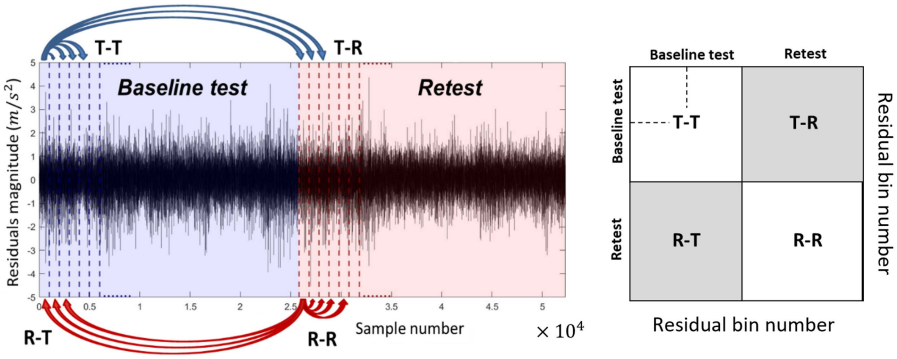

Fig. 3. Illustration of the data bin comparisons. On the left, the dotted lines represent the divisions of the residuals bins to be compared against the remainders. The arrows represent the two-sample comparisons. On the right, the comparison map is shown. This will then allow the user to visualise the location of the inconsistencies in the residual signals.

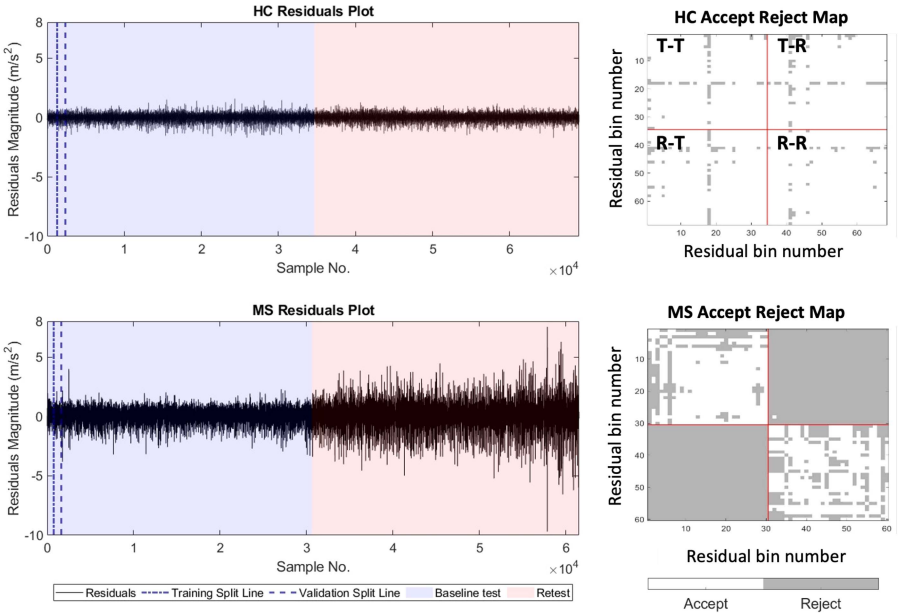

Fig. 4. Left: Typical residuals patterns for a HC (top) and an MS (bottom) individual. Right: Examples of the corresponding accept-reject maps. Here, the red lines mark the T-T, T-R, R-T and R-R quadrants. Gait inconsistencies in the form of null hypothesis rejections are flagged by the grey squares across all bin comparisons.

points, which are referred to as data bins in the following paragraphs. In the context of the dataset used in this work, which contains test and retest data, the MMD hypothesis test was employed to compare all the residuals bins from baseline test vs. baseline test (T-T), baseline test vs. retest (T-R), retest vs. baseline test (R-T) and retest vs. retest (R-R). This procedure, which is depicted in Fig. 3 led to the creation of an *accept-reject map*, which can be visually interpreted using the examples provided in Fig. 4.

Upon visual inspection of residual signals corresponding to the MS group, it was noticed that, in several cases, the residual variance is increasing towards the end of a walking test, which might be an indication of fatigue over prolonged periods of walking. Thus, to compare the consistency of the gait patterns between the two tests, and to capture the seemingly fatiguing behaviour described previously, it was then decided to optimize the bandwidth of the kernel by minimizing the accuracy of the test in the area of the map enclosed by the dotted lines in the T-T quadrant, (i.e., using the first half of the baseline residuals, as seen in Fig. 3), assuming that a person can walk relatively consistently during half of the baseline test. For clarification, this optimization procedure was employed across both groups of subjects.

More formally, the optimal bandwidth with an L2 regularization imposed was calculated as:

$$\sigma_{optim=} \arg\min_{\sigma}(accuracy(\sigma) + \sigma^2) \tag{6}$$

To arrive at the optimal kernel bandwidth for each subject, it is important to ensure the robustness of the solution in a couple of ways. Since typical optimization schemes involving gradient descend are not feasible here due to the bootstrapping

procedure involved in the hypothesis testing, a gradient-free optimization method was chosen. To ensure convergence to a global minimum, multiple runs of *fmincon* interior-point and *fminbnd* MATLAB-built in algorithms have been used, with a search interval restricted to [0.0001, 10] and a cost function value tolerance set to 1e-5. For the extensive descriptions of the optimization algorithms used, the reader is referred to [45], [46], [47] or [48], [49] for *fmincon* and *fminbnd* respectively. Finally, the value of the bandwidth corresponding to the minimum function value of all runs was taken as the optimal value.

Once the optimal bandwidth is found, a counting task follows. The MMD-hypothesis test is implemented, by comparing all the residual bins as described previously, and the *accept-reject map* is populated with the results returned by the hypothesis test as a yes/no survey of whether the null hypothesis is accepted for each of the comparisons. Finally, the percentage of null hypothesis rejections is computed in each of the quadrants (T-T, T-R, R-T and R-R) as a measure of gait consistency. A higher percentage signifies a higher number of gait anomalies found in the residual patterns and therefore a less consistent gait. This final counting step concludes the workflow of the proposed methodology for the objective quantification of gait consistency.

V. Demonstration on the Chosen Case Study

The effectiveness of this newly proposed approach, comprising of utilizing ARX modelling and MMD-hypothesis testing, for determining gait consistency is evaluated in this study by applying it to the dataset discussed in Section III. The objectives targeted here are (1) to verify whether the consistency of gait is altered by the presence of a locomotor disease (i.e., MS in the case of this work) and (2) quantify the effect of variations in testing conditions.

To begin with, ARX model coefficients were computed on the training set (consisting of the first straight line walking bout of the baseline test), while the model order has been established on the validation set (using the second straight line walking bout of the baseline test). Then, once the model order and coefficients have been established, the model residuals were computed across the entire baseline test and retest. Examples of the residual patterns for a HC and an individual with MS, when the retest has been performed one hour apart, can be seen in Fig. 4. Here, the baseline test is highlighted by the blue region, while the red region marks the retest. Moreover, the training and validation regions for constructing the ARX models are marked by the blue dotted lines at the beginning of the baseline test. Upon visual inspection, qualitatively, it can be seen that this HC is able to maintain a stable gait throughout both tests, as the variance of the residuals remains relatively constant. The individual affected by MS, on the other hand, displays an inconsistent gait during the retest, when compared to the baseline test. An increase in the variance of the residuals is also noted towards the end of the retest. Next, the corresponding *accept-reject maps* can also be seen on the right in Fig. 4. These were created by cross comparing the residual bins using the MMD-hypothesis test, for which the number of permutations was set to 500, and the confidence level was set to 99%. Firstly, the HC *accept-reject map* in Fig. 4 shows the clear consistent gait pattern of the HC subject, as only a few grey squares are visible, indicating which residual bins were different from the rest, as the null hypothesis was rejected. In contrast, the accept-reject map for the individual affected by MS confirms the qualitative findings in the corresponding residual plot, as the grey areas in the T-R and R-T quadrants further reinforce the gait inconsistencies between the two repeated tests.

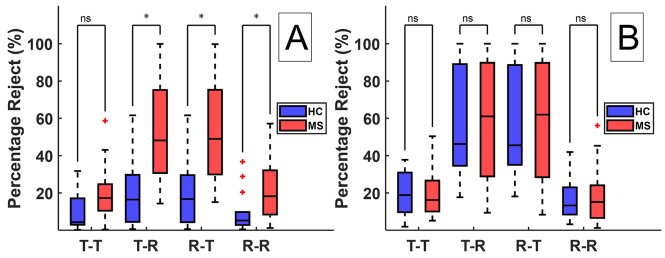

Fig. 5. Summary of the results. (a) - one-hour apart group, (b) - one-week apart group. The horizontal line inside the boxes represents the median value, with the box showing the interquartile range. The whiskers indicate the 2.7 standard deviations range, considering a Gaussian distribution. The outlying data is shown using red crosses. * represents statistical significant difference, ns represents a non-significant result.

Once the test-retest residual sequences and the *accept-reject maps* have been computed for all the participants in this case study, further statistical analysis was required to accomplish the two rather exploratory objectives of this study. Therefore, the statistical comparison of the percentages of null hypothesis rejections as a measure of gait anomalies in all the quadrants of the *accept-reject map* was performed using the Mann-Whitney U-test with a minimum significance alpha level of 1.25%, following Bonferroni correction, which accounts for the multiple comparisons ($\alpha^* = 0.05$ / 4 comparisons). Type II error was evaluated using Cohen's d estimate, where values of 0.1, 0.3 and 0.5 were used as thresholds for small, medium, and large effect sizes, according to [50]. All statistical tests discussed in this paragraph were performed in MATLAB 2021b (MathWorks, Inc., Natick, MA, USA).

The comparison between the percentages of null hypothesis rejections in all the quadrants of the *accept-reject maps* are shown in Fig. 5. The statistical comparison between the HC and MS groups who completed the retest one hour apart is shown in Fig. 5(a), while Fig. 5(b) shows the comparison for those who completed the retest one week apart.

Analysing Fig. 5(a), as indicated by the T-R and R-T comparisons, significant differences in the consistency of gait patterns of MS subjects were found even when the retest was performed one hour apart ($p = 0.0002$ and $p = 0.0003$ for T-R and R-T respectively). Although not significant ($p = 0.0208$), the T-T comparison suggests that the MS group displayed a higher number of gait anomalies during the baseline test. This is also indicated by the large effect size ($d = 0.77$) recorded for this comparison and might be interpreted as an indication of fatigue or balance and coordination difficulties during prolonged periods of gait. On the other hand, a statistically significant higher number of gait inconsistencies was recorded during the retest for the MS group ($p = 0.0048$), as indicated by the R-R

TABLE II
DESCRIPTIVE STATISTICS FOR THE INVESTIGATED COMPARISONS, TOGETHER WITH P-VALUES FOR THE INDEPENDENT MANN-WHITNEY U TEST WITH BONFERRONI CORRECTION AND ASSOCIATED EFFECT SIZES

	HC (A)	MS (A)	HC (B)	MS (B)	HC vs. MS (A)	HC vs. MS (B)
	Median (min, 25^th percentile, 75^th percentile, max)				p-value (d)	
T-T (%)	4.33 (0.17, 2.98, 16.62, 31.74)	17.25 (0.39, 10.58, 24.11, 58.69)	18.83 (1.90, 10.21, 30.01, 37.78)	16.20 (5.10, 10.21, 24.25, 50.38)	0.0208 (0.77)	0.8694 (0.10)
T-R (%)	16.44 (0.57, 5.09, 27.39, 61.62)	48.14 (14.33, 30.91, 75.03, 99.90)	46.23 (17.69, 35.96, 88.13, 100)	61.06 (9.38, 30.20, 89.44, 100)	**0.0002** (1.33)	1 (0.07)
R-T (%)	16.73 (0.57, 4.88, 27.48, 61.62)	48.98 (15.06, 30.21, 75.18, 99.79)	45.54 (18.11, 36.17, 87.88, 100)	61.98 (8.33, 29.90, 89.35, 100)	**0.0003** (1.32)	0.9812 (0.07)
R-R (%)	5.20 (0.48, 2.82, 9.09, 36.73)	18.20 (1.21, 8.66, 31.76, 57.18)	13.28 (3.17, 8.49, 22.95, 41.94)	15.19 (1.22, 6.69, 23.44, 56.12)	**0.0048** (0.84)	0.7070 (0.12)

$kp - value < 0.05$ (k = number of multiple comparisons, equal to 4) are in bold.
A: Group 1, who performed the retest one hour apart; B: Group 2, who performed the retest one week apart.

comparison, while the HC group showed a variance decrease. It should be noted that although, the T-R and R-T comparisons appear identical, they are not, due to the bootstrapping procedure involved in the MMD hypothesis test. Large effect sizes were recorded across all comparisons for the first group.

Analysing Fig. 5(b), no significant differences between the HC and MS groups were found across all comparisons. Interestingly, the within-test comparisons (T-T and R-R) do not show a difference between HCs and MS, in contrast to the differences found within group A. Moreover, a small effect size was recorded across all comparisons. The associated descriptive statistics characterising the boxplots are presented in Table II.

VI. DISCUSSION

ARX residual modelling has been put forward as a new way of attempting to reveal gait inconsistencies for both healthy and pathological individuals. The idea of monitoring the residuals can immediately highlight departures from the normal stable gait, as a change in the residual pattern of a dynamic model of gait. This difference occurs when the previously identified ARX model is not able to make a good prediction, due to significant changes in the system's dynamics. Interestingly, although the model orders and coefficients are individually selected for each participant, once computed at the beginning of the baseline assessment, they do not have to be recomputed again. This advantage proved to be extremely useful at highlighting the influence of the confounding factors present at the follow-up assessment. It should also be noted, that given the uniqueness of the gait patterns, the models are not transferrable between individuals, as a result of significant differences in the number of lags and coefficients. Furthermore, it was observed that residual patterns obtained during the train and validation walking-bouts were not affected by the presence of severely impaired movement, slow walking, asymmetry, compensatory movements, or the utilization of walking aids. Inconsistencies were observed at later points during the baseline assessment or the one-week apart retest. Additionally, another inherent advantage of employing the proposed approach is that it avoids the need for gait event identification algorithms, which can be susceptible to inaccuracies due to the aforementioned factors that impact their accuracy.

After successfully implementing the residual modelling task, quantifying gait consistency through evaluating the similarity of residual sequences at various time points was approached by utilising non-parametric statistical hypothesis testing. For this task, a non-parametric hypothesis test is a necessity. Therefore, in the second part of the proposed methodology, this article introduced the MMD-hypothesis test which offers a kernel-embedding of the distribution of the residuals, being able to account for all the information present in the distributions. This is advantageous because end users are not required to specify a-priori the particular features of residual distributions that the statistical test should detect. Instead, the kernel trick allows the user to effectively assess infinite statistical moments through the use of inner products in a feature space [33]. Finally, this article also presented the *accept-reject maps* as a way of objectively quantifying gait consistency. The idea of monitoring the ARX model residuals is fundamentally connected to the requirement of a hypothesis test, as the two parts of this newly proposed methodology can only exist in conjunction. Next, it is perhaps important to revisit and further discuss the meaning of the results presented in Section V.

The development of this methodology enabled the first objective to be addressed. This objective aimed to assess whether the presence of a disease alters the consistency of gait. Here, the disease in question was MS, which is known to manifest through gait balance and coordination deficits. Clearly, Fig. 5(a) highlights that even when the retest is performed one-hour apart, and all the external factors were controlled (i.e., the sensors were not removed and repositioned between the two repeated recordings, the subjects wore the same shoes, and enough time was allowed for resting in between the two tests), the individuals affected by MS have difficulties maintaining a stable and consistent gait. Because of this, the proposed methodology might have the potential of being a viable tool for verifying the effect of short-term clinical interventions, such as the Remote Ischaemic Preconditioning (RIPC) [51]. In addition to this, the increased sensitivity of the method yields it suitable for quantifying the within-test gait consistency in pathological populations, yielding

an overall consistency metric, given that less data is used for the kernel bandwidth optimization task.

The second objective examined whether gait consistency affected is influenced by alterations in testing conditions. The results presented in Fig. 5(b) clearly highlight that further work is required to generalize the models and remove the influence of the confounding factors. In this case, their influence seems to be greater than the influence of the disease itself, which might mask the evolution of the gait patterns at follow up assessments. Here, the one-week interval between tests enabled the proposed methodology to be evaluated under more realistic follow-up assessment scenarios, where changes in sensor placement, differences in the time of assessment, prior physical activity before testing, differences in subject's footwear, and other factors might occur. To clarify, the included MS subject did not undergo any disease-related therapeutic interventions and the one-week apart was deliberately chosen to ensure a stable disease status during the study period. Therefore, by controlling these factors, the effects of disease progression or treatment effects were isolated. The findings of the study revealed statistically significant results for the healthy population when comparing the baseline assessment with the one-week apart retest. These results highlighted the presence of natural variability in gait patterns. If the modelling approach had demonstrated sufficient generality, the authors would have expected no significant differences for the T-R and R-T comparisons using group B. Such outcomes would have indicated successful isolation of the natural variability. Remarkably, the same problem is well known in the SHM field [52], [53], [54], [55], which is where the inspiration for this modelling approach comes from. This highlights the fact that the confounding factors arose from environmental changes in testing conditions are detrimental to the assessment of the condition of the system being analysed, regardless of the nature of the system (i.e., being a person, a bridge, a plane etc.). To mitigate these problems, various tools were developed in SHM, such as cointegration [54], [55], and seem worthy of investigation for the future work.

Because of the nature of the analysis, the findings of this case-study cannot be directly compared to any results already presented in the relevant literature. While the introduced concepts are fundamentally different, perhaps the closest resembling study is the work of Angelini et al. in [7], who assessed the between-session reliability of several temporal, variability and balance gait metrics using data collected one-week apart from the same group (B) of HCs and patients affected by MS. The study investigated between-session gait metrics reliability, which refers to how consistent the included gait metrics remain over time. Good to excellent agreement between the two repeated tests was reported. In contrast, the present study directly investigated the consistency of the gait patterns themselves and found significant inconsistencies between test-retest gait patterns. Although the same group of subjects was used for the one-week apart comparison, this outcome was expected, due to the increased sensitivity offered by the ARX-based method and the MMD-based hypothesis test. In addition to this, several other studies attempted to quantify the reliability of several spatio-temporal metrics for subjects affected by MS. Morris et al. [56] quantified gait consistency during over a five-hour interval, from morning to afternoon in MS patients, and provided a comparison to HCs. The study revealed that even though the gait metrics obtained for the patients with MS were different when compared to those extracted from HCs, they remained consistent throughout the monitoring window, in contrast to the results of the short-term comparison presented in this study. However, the analysis did not include a comprehensive set of gait metrics, as only the gait speed, cadence, stride length and double limb support percentage were examined. In addition to this, the study was based on the 10-meter walking test, which is known to suffer from lack of precision [57]. Another study utilising test-retest data was the attempt of Berg-Hansen et al. [24] to measure the potential effects of rehabilitation for MS subjects. Significant differences from test to retest were found for all spatio-temporal parameters included in the study. Yet, because no control group was included in this study, the presented outcome cannot be attributed to the active rehabilitation and cannot be either attributed to varying testing conditions. On the other hand, our case-study included the control group and has shown that gait inconsistencies have the potential of serving as an indicator of the presence of MS, when the environmental testing conditions are maintained constant.

It is also worth discussing the possible influence of walking aids and the presence of asymmetry on the results, as the latter is often considered an indicator of MS [20]. Here, the novel data-driven approach utilized in this study extends beyond the investigation of normal gait patterns and can effectively analyse complex pathological gait, regardless of its severity or the use of walking aids. To test the robustness of the proposed methodology, 18 individuals affected by MS who relied on walking aids were also included. Consequently, our findings indicated that the presence of walking aids did not negatively affect the residual modelling task. This suggests that employing an adequate number of lags to comprehensively capture the dynamic relationship between the lower limbs and the upper body ensured that the residuals resembled white noise without discernible patterns. Moreover, it is also important to note that the ARX modelling procedure presented here inherently deals with situations where asymmetry between the lower limbs is present, i.e. instances characterised by temporal differences, variations in signal amplitude, or noisier acceleration signals recorded on one leg in comparison to the other. This aspect is being handled by treating the left and right limbs as separate entities and assigning a varying number of lags and unique coefficients, as needed. The model parameters are automatically tailored to accommodate asymmetry effectively, using the BIC, as explained in Section IV. Furthermore, is it also worth noting that the methodology presented in this work is not intended to replace traditional gait analysis, but rather to serve as a valuable augmentation to the clinical assessment of gait consistency.

Having stressed the advantages and potential uses of the newly proposed methodology for gait consistency quantification, some though must also be given to the possible limitations. The first obvious limitation concerns the subject specific ARX models, since no significant within-test differences were found between

HCs and MS, except for the RR comparison in group A. There are a few reasons for these results. Firstly, if the model learns the already impaired gait pattern of an individual with MS and if that particular subject sustains the same impaired gait pattern throughout the entire walking test, the residuals will still resemble white noise and have a low variance. Secondly, the bandwidth optimization was performed using the first half of the baseline test data. While this approach is ideal for verifying the gait consistency between repeated tests, it might mask important deviations from the norm of the baseline data. To counteract this problem, a feasible alternative would be to use a smaller proportion of the baseline data for the optimization of the kernel bandwidth. Moreover, it is also important to acknowledge that although the study utilized raw acceleration data, in future applications where device-agnostic methodologies are desired, filtering techniques could be implemented prior to the ARX modelling task. However, in this particular case, such a requirement was unnecessary as the same devices were consistently employed throughout the data acquisition process.

For the second part of the methodology, a potential disadvantage of the MMD-based hypothesis test is the computational overhead. Yet, this was reduced by avoiding the quadratic-time computational cost in the permutation loop. Moreover, even though the T-R and R-T comparisons may not always be the same (due to the bootstrapping procedure involved in creating the two artificially symmetric distributions), the differences are negligible, and the computational time can be further reduced in half, if only considering the upper or lower diagonal matrices of the *accept-reject map*. A possible improvement would be to use the updated versions of the MMD, such as its linear time estimate [44] or the B-tests [58].

VII. Conclusion

The present article has introduced a novel methodology for objectively quantifying gait consistency for both healthy and pathological individuals. While clinical conclusions are probably not advisable, instead, it is perhaps more important to discuss the main ideas introduced by this article and their future usages. The idea of a data-based modelling approach, in the form of ARX residual modelling, has been applied in the context of gait analysis. The aim has been to investigate whether monitoring the residuals can lead to insight into, or enhancement of, the understanding of gait consistency in both healthy and pathological populations. Thus, once an ARX model has been established, good predictions can only be obtained providing a stable and controlled gait, similar to the one displayed during the learning phase. As a result, the residuals should have a constant variance. On the contrary, in the presence of gait inconsistencies, an obvious departure from the constant residual variance should be recorded. This modelling approach immediately lends itself for monitoring the consistency of gait during clinical walking assessments. However, obtaining an objective measure of gait consistency is only possible if the residual modelling is used in conjunction with statistical hypothesis testing. To this end, the MMD-based hypothesis test has been introduced, offering enhanced sensitivity to gait inconsistencies. Finally, by considering smaller data segments, an objective measure of consistency has been provided, by cross comparing all the smaller data segments.

The result of the most immediate importance for gait analysis community is that this newly proposed methodology revealed the detrimental effects of varying assessment conditions on gait pattern consistency. Therefore, the obvious direction of the future work targets the elimination of environmental variations, which will then allow the long-term monitoring of gait progression in longitudinal studies.

Acknowledgment

We would like to acknowledge Ellen Buckley, Lorenza Angelini, Siva Nair and David Paling for providing the data used in this study, which have been collected within the NIHR Sheffield Clinical Research Facility (CRF) as part of NIHR Sheffield Biomedical Research Centre (BRC) activities. We would like to thank all participants for giving their time to support this research.

References

[1] D. A. Winter, *Biomechanics and Motor Control of Human Movement*, 4th ed. Hoboken, NJ, USA:Wiley, 2009.
[2] W. Zijlstra and A. L. Hof, "Displacement of the pelvis during human walking: Experimental data and model predictions," *Gait Posture*, vol. 6, no. 3, pp. 249–262, 1997.
[3] A. Polhemus et al., "Walking on common ground : A cross-disciplinary scoping review on the clinical utility of digital mobility outcomes," *npj Digit. Med.*, vol. 4, 2021, Art. no. 149.
[4] A. D. Kuo, "The six determinants of gait and the inverted pendulum analogy: A dynamic walking perspective," *Hum. Movement Sci.*, vol. 26, no. 4, pp. 617–656, 2007.
[5] F. Pecoraro et al., "Assessment of level-walking aperiodicity," *J. NeuroEng. Rehabil.*, vol. 3, pp. 1–9, 2006.
[6] J. Chakraborty and A. Nandy, "Periodicity detection of quasi-periodic slow-speed gait signal using imu sensor," in *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body and Motion*. Berlin, Germany: Springer Int. Publishing, 2019, pp. 140–152.
[7] L. Angelini et al., "Wearable sensors can reliably quantify gait alterations associated with disability in people with progressive multiple sclerosis in a clinical setting," *J. Neurol.*, vol. 267, no. 10, pp. 2897–2909, 2020.
[8] L. Ljung, *System Identification, Theory for the User*. Englewood Cliffs, NJ, USA: Prentice Hal PTR, 1999.
[9] A. Gretton et al., "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 25, pp. 723–773, 2012.
[10] F. Kluge et al., "Towards mobile gait analysis: Concurrent validity and test-retest reliability of an inertial measurement system for the assessment of spatio-temporal gait parameters," *Sensors*, vol. 17, no. 7, 2017, Art. no. 1522.
[11] J. J. Craig et al., "Instrumented balance and walking assessments in persons with multiple sclerosis show strong test-retest reliability," *J. NeuroEngineering Rehabil.*, vol. 14, no. 1, 2017, Art. no. 43.
[12] P. Decavel, T. Moulin, and Y. Sagawa, "Gait tests in multiple sclerosis: Reliability and cut-off values," *Gait Posture*, vol. 67, pp. 37–42, 2019.
[13] D. M. Urquhart, M. E. Morris, and R. Iansek, "Gait consistency over a 7-day interval in people with Parkinson's disease," *Arch. Phys. Med. Rehabil.*, vol. 80, no. 6, pp. 696–701, 1999.
[14] C. Buckley et al., "Upper body accelerations as a biomarker of gait impairment in the early stages of Parkinson's disease," *Gait Posture*, vol. 71, pp. 289–295, 2019.
[15] A. Sidiropoulos, R. Magill, and A. Gordon, "Coordination of the upper and lower extremities during walking in children with cerebral palsy," *Gait Posture*, vol. 86, pp. 251–255, 2021.
[16] M. G. Pandy, Y.-c. Lin, and H. J. Kim, "Muscle coordination of mediolateral balance in normal walking," *J. Biomech.*, vol. 43, no. 11, pp. 2055–2064, 2010.
[17] J. L. Allen and R. R. Neptune, "Three-dimensional modular control of human walking," *J. Biomech.*, vol. 45, no. 12, pp. 2157–2163, 2012.

[18] J. M. Hausdorff, D. A. Rios, and H. K. Edelberg, "Gait variability and fall risk in community-living older adults: A. 1-year prospective study," *Arch. Phys. Med. Rehabil.*, vol. 82, no. 8, pp. 1050–1056, 2001.

[19] S. B. Richmond et al., "A temporal analysis of bilateral gait coordination in people with multiple sclerosis," *Mult. Scler. Related Disord.*, vol. 45, 2020, Art. no. 102445.

[20] L. Angelini et al., "A multifactorial model of multiple sclerosis gait and its changes across different disability levels," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 11, pp. 3196–3204, Nov. 2021.

[21] R. I. Spain et al., "Body-worn motion sensors detect balance and gait deficits in people with multiple sclerosis who have normal walking speed," *Gait Posture*, vol. 35, no. 4, pp. 573–578, 2012.

[22] M. Pau et al., "Clinical assessment of gait in individuals with multiple sclerosis using wearable inertial sensors: Comparison with patient-based measure," *Mult. Scler. Related Disord.*, vol. 10, pp. 187–191, 2016.

[23] Y. Mao et al., "Estimation of stride-by-stride spatial gait parameters using inertial measurement unit attached to the shank with inverted pendulum model," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 1391.

[24] P. Berg-Hansen et al., "Sensor-based gait analyses of the six-minute walk test identify qualitative improvement in gait parameters of people with multiple sclerosis after rehabilitation," *J. Neurol.*, vol. 1, pp. 3723–3734, 2022.

[25] G. A. Cavagna, H. Thys, and A. Zamboni, "The sources of external work in level walking and running," *J. Physiol.*, vol. 262, no. 3, pp. 639–657, 1976.

[26] F. Rasouli and K. B. Reed, "Identical limb dynamics for unilateral impairments through biomechanical equivalence," *Symmetry*, vol. 13, no. 4, 2021, Art. no. 705.

[27] C. R. Farrar and K. Worden, "An introduction to structural health monitoring," *Phil. Trans. R. Soc. A*, vol. 365, no. 1851, pp. 303–315, 2007.

[28] S. D. Silva, M. Dias, and V. Lopes, "Damage detection in a benchmark structure using AR-ARX models and statistical pattern recognition," *J. Braz. Soc. Mech. Sci. Eng.*, vol. 29, no. 2, pp. 174–184, 2007.

[29] H. Sohn et al., "Structural health monitoring using statistical pattern recognition techniques," *ASME. J. Dyn. Sys., Meas., Control.*, vol. 123, no. 4, pp. 706–711, 2001.

[30] H. Sohn, J. A. Czarnecki, and C. R. Farrar, "Structural health monitoring using statistical process control," *J. Struct. Eng.*, vol. 126, no. 11, pp. 1356–1363, 2000.

[31] A. Entezami, H. Shariatmadar, and S. Mariani, "Early damage assessment in large-scale structures by innovative statistical pattern recognition methods based on time series modeling and novelty detection," *Adv. Eng. Softw.*, vol. 150, 2020, Art. no. 102923.

[32] P. E. Carden and J. M. W. Brownjohn, "ARMA modelled time-series classification for structural health monitoring of civil infrastructure," *Mech. Syst. Signal Process.*, vol. 22, no. 2, pp. 295–314, 2008.

[33] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.

[34] B. K. Sriperumbudur et al., "Hilbert space embeddings and metrics on probability measures," *J. Mach. Learn. Res.*, vol. 11, pp. 1517–1561, 2009.

[35] L. Comber, R. Galvin, and S. Coote, "Gait deficits in people with multiple sclerosis: A systematic review and meta-analysis," *Gait Posture*, vol. 51, pp. 25–35, 2017.

[36] J. F. Kurtzke, "Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS)," *Neurology*, vol. 33, no. 11, pp. 1444–1452, 1983.

[37] G. E. P. Box et al., *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.

[38] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[39] P. Stoica and Y. Selén, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[40] K. Fukumizu et al., "Kernel measures of conditional dependence," in *Proc. 20th Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 489–496.

[41] J. R. Lloyd and Z. Ghahramani, "Statistical model criticism using kernel two sample tests," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 829–837.

[42] A. Gretton et al., "A kernel statistical test of independence," in *Proc. 20th Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 585–592.

[43] D. Garreau, W. Jitkrittum, and M. Kanagawa, "Large sample analysis of the median heuristic," 2018.

[44] A. Gretton et al., "Optimal kernel choice for large-scale two-sample tests," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1205–1213.

[45] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Math. Program.*, vol. 89, pp. 149–185, 2000.

[46] R. A. Waltz et al., "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Math. Program.*, vol. 107, pp. 391–408, 2006.

[47] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. Optim.*, vol. 9, no. 4, pp. 877–900, 2006.

[48] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations*. vol. 59. Englewood Cliffs, NJ, USA: Prentice Hall PTR, 1979.

[49] R. P. Brent, *Algorithms for Minimization Without Derivatives*, 1st ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1973.

[50] J. Cohen, "The significance of a product moment rs," in *Statistical Power Analysis for the Behavioral Sciences*. Cambridge, MA, USA: Academic Press, 1977, ch. 3, pp. 75–107.

[51] C. Chotiyarnwong et al., "Effect of remote ischaemic preconditioning on walking in people with multiple sclerosis: Double-blind randomised controlled trial," *BMJ Neurol. Open*, vol. 2, no. 1, 2020, Art. no. e249.

[52] H. Sohn, K. Worden, and C. R. Farrar, "Statistical damage classification under changing environmental and operational conditions," *J. Intell. Mater. Syst. Structures*, vol. 13, no. 9, pp. 561–574, 2002.

[53] H. Sohn, "Effects of environmental and operational variability on structural health monitoring," *Phil. Trans. R. Soc. A*, vol. 365, no. 1851, pp. 539–560, 2006.

[54] E. J. Cross, K. Worden, and Q. Chen, "Cointegration: A novel approach for the removal of environmental trends in structural health monitoring data," *Proc. R. Soc. A*, vol. 467, pp. 2712–2732, 2011.

[55] K. Worden et al., "A multiresolution approach to cointegration for enhanced SHM of structures under varying conditions–An exploratory study," *Mech. Syst. Signal Process.*, vol. 47, no. 1, pp. 243–262, 2014.

[56] M. E. Morris et al., "Changes in gait and fatigue from morning to afternoon in people with multiple sclerosis," *J. Neurol. Neurosurgery Psychiatry*, vol. 72, no. 3, pp. 361–365, 2002.

[57] P. Feys et al., "Within-day variability on short and long walking tests in persons with multiple sclerosis," *J. Neurological Sci.*, vol. 338, no. 1, pp. 183–187, 2014.

[58] W. Zaremba, A. Gretton, and M. Blaschko, "B-tests: Low variance kernel two-sample tests," *Proc. Adv. Neural Inf. Process*, 2013, pp. 755–763.