Article:
Stihi, A., Mazzà, C., Cross, E. et al. (2025) Hierarchical Gaussian processes for characterizing gait variability in multiple sclerosis. Data-Centric Engineering, 6. e36. ISSN: 2632-6736

https://doi.org/10.1017/dce.2025.10009

Data-Centric Engineering (2025), 6: e36
doi:10.1017/dce.2025.10009

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Hierarchical Gaussian processes for characterizing gait variability in multiple sclerosis

Alexandru Stihi[1,2], Claudia Mazzà[1], Elizabeth Cross[2] and Timothy James Rogers[2]

[1]Department of Mechanical Engineering, Insigneo Institute for In Silico Medicine, University of Sheffield, Sheffield, UK
[2]Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Sheffield, UK
Corresponding author: Timothy James Rogers; Email: tim.rogers@sheffield.ac.uk

Received: 27 June 2024; **Revised:** 25 March 2025; **Accepted:** 26 March 2025

Keywords: gait analysis; Gaussian process; heteroscedastic; hierarchical; multiple sclerosis

Abstract

Reduction in mobility due to gait impairment is a critical consequence of diseases affecting the neuromusculoskeletal system, making detecting anomalies in a person's gait a key area of interest. This challenge is compounded by within-subject and between-subject variability, further emphasized in individuals with multiple sclerosis (MS), where gait patterns exhibit significant heterogeneity. This study introduces a novel perspective on modeling kinematic gait patterns, recognizing the inherent hierarchical structure of the data, which is gathered from contralateral limbs, individuals, and groups of individuals comprising a population, using wearable sensors. Rather than summarizing features, this approach models the entire gait cycle functionally, including its variation. A Hierarchical Variational Sparse Heteroscedastic Gaussian Process was used to model the shank angular velocity across 28 MS and 28 healthy individuals. The utility of this methodology was underscored by its granular analysis capabilities. This facilitated a range of quantifiable comparisons, spanning from group-level assessments to patient-specific analyses, addressing the complexity of pathological gait patterns and offering a robust methodology for kinematic pattern characterization for large datasets. The group-level analysis highlighted notable differences during the swing phase and towards the end of the stance phase, aligning with previously established literature findings. Moreover, the study identified the heteroscedastic gait pattern variability as a distinguishing feature of MS gait. Additionally, a novel approach for lower limb gait asymmetry quantification has been proposed. The use of probabilistic hierarchical modeling facilitated a better understanding of the impaired gait pattern, while also expressing potential for extrapolation to other pathological conditions affecting gait.

Impact Statement

Quantifying gait patterns can enhance the understanding of disease progression or response to interventions, which is particularly relevant for individuals with multiple sclerosis (MS). In this context, wearable sensors emerge as a cost-effective and unobtrusive tool for patient monitoring. However, the quantitative evaluation of the impact of a disease on gait and mobility using data collected by wearable sensors remains a challenging task. This is primarily due to substantial variability between individuals and temporal fluctuations within a single individual's data. The work presented here proposes a novel data-driven probabilistic approach which can help reveal which part of the gait cycle is most affected by a disease using a Bayesian machine learning technique, the Hierarchical Variational Sparse Heteroscedastic Gaussian Process.

1. Introduction

Gait is a complex dynamic process that has received increased attention by the scientific community due to its relevance in understanding human health and pathology. Gait impairment, often considered a hallmark of neuromuscular diseases (Polhemus et al., 2021; Cicirelli et al., 2022), is notably exemplified in multiple sclerosis (MS). This neurodegenerative condition is characterized by the inflammatory-mediated demyelination of axons within the central nervous system (Comber et al., 2017). For an enhanced understanding and quantification of the disease, it is necessary to accurately characterize the lower limb distal motion, as it is often affected as a result of alterations in distal muscle involvement (Filli et al., 2018; Pau et al., 2021; Polhemus et al., 2021). As such, the shank angular velocity emerges as a potential signal of interest within this context. While the existing literature predominantly employs the shank angular velocity for gait cycle detection, emphasizing its effectiveness in identifying key gait event landmarks (Pacini Panebianco et al., 2018), the authors assert that the full potential of this signal remains largely unexplored. Additionally, advancements in wearable sensor technology, particularly inertial measurement units (IMUs), have led to their increased application in clinical gait analysis (Angelini et al., 2020). Owing to their flexibility, IMUs present practical solutions for quantifying lower limb distal motion, providing a direct measure of the shank angular velocity, and further motivating the clinical utility of this signal. In light of these considerations, the current work proposes a new approach which seeks to model the full kinematic signal of the shank angular velocity using a data-driven approach. As a first step before exploring longitudinal gait changes, this model is used to reveal regions of the gait cycle that are most affected by the disease or exhibit the greatest variation between contralateral limbs or individuals.

A gait cycle—which consists of the *stance phase* (accounting for approximately the first 60% of the gait cycle) and the *swing phase* (accounting for the remainder of the gait cycle)—ends and begins with the *heel strike* event (Rueterbories et al., 2010). During the stance phase, the limb bearing the body weight transitions from heel strike to *toe-off*. During the *mid-stance*, the body is transitioned forward, and the opposite, contralateral limb is in the swing phase. This part of the gait cycle is characterized by a small base of support and a relatively high center of gravity, making the walker least stable. The swing phase, encompassing *initial swing*, the *mid-swing*, and *terminal swing*, propels the limb forward in preparation for the subsequent heel strike. Figure 1 shows typical examples of shank angular velocity signals during a

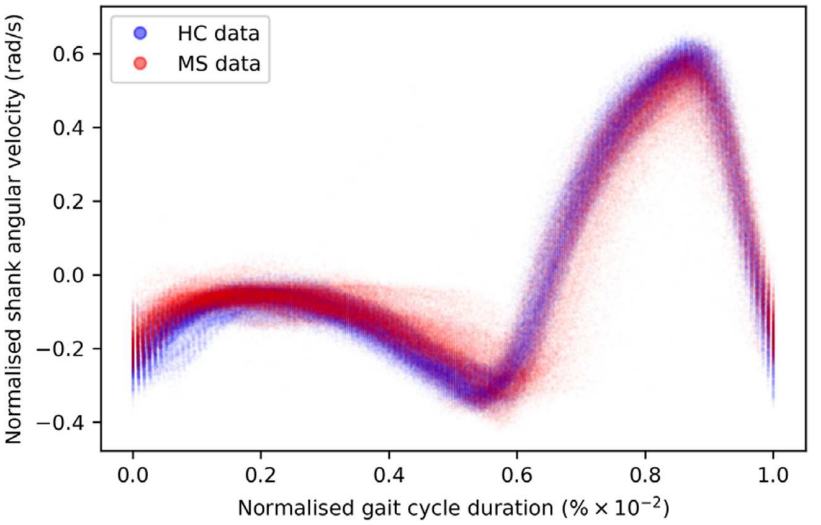

Figure 1. *Comparison of shank angular velocity data between healthy controls (HC) and individuals with multiple sclerosis (MS). The figure presents an aggregate of data points collected using inertial measurement units from both left and right limbs, encompassing 7899 gait cycles from 28 healthy controls and 7105 gait cycles from 28 individuals affected by multiple sclerosis.*

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

gait cycle, collected using wearable sensors, for two distinct groups of individuals: healthy controls (HCs) and people with MS (PwMS). Relative to the HC group, the MS group exhibits increased gait pattern variability, particularly discernible from *mid-stance* to *mid-swing*. For clarity, the term "variability" is commonly employed in gait analysis to signify stride-to-stride fluctuations during walking, often serving as an indicator of gait impairment (Moon et al., 2016). However, within the scope of this study, "variability" refers to the dispersion of shank angular velocity around its mean characteristic pattern throughout the normalized gait cycle. Although a direct comparison with the relevant literature presents challenges, as the majority of the studies focus on joint kinematics, rather than segment kinematics (as in the case of the present study), similar trends have been previously documented. Crenshaw et al. (2006) reported a significantly higher knee and ankle joint angle variability for PwMS. Kelleher et al. (2010) proposed that individuals with MS experience insufficient propulsion from the ankle plantar flexor muscles and lack fine motor control during the swing phase, perhaps as a result of favoring the more proximal muscle groups. A reduced range of ankle flexion was also confirmed by Nogueira et al. (2013), even for PwMS in the prodromal phase of the disease. Similar trends were also reported by Severini et al. (2017). More recently, Salehi et al. (2020) investigated the deviation phase as a measure of coordination variability and reported significant differences between HC and MS gait during the stance and swing phases. The increased variability observed in the MS group may also be attributed to inefficient gait compensations prompted by factors such as muscle weakness, spasticity, fatigue, or balance impairments (Nogueira et al., 2013; Socie et al., 2013; Gil-Castillo et al., 2020).

From a modeling perspective, it can be seen that the gait signals are exhibiting a number of interesting features. First, the relationship between the input domain and the shank angular velocity is not linear. Second, the variance in the data is not constant across the input space. This phenomenon is known as a heteroscedastic noise process, where the variance in the data is dependent on the input (Or the noise variance changes across the input domain and can be modeled as a function of the input.). In contrast, the process when the variance is independent of the input is referred to as a homoscedastic noise process. Third, inspecting Figure 1, which displays the datapoints from repeated gait cycles collected during straight-line walking for multiple individuals belonging to both groups, it is clear that there is a shared underlying pattern. This observation not only underscores the commonalities in gait dynamics between HC and PwMS but also prompts an exploration of the hierarchical structure inherent in the process of data acquisition during gait assessments. Starting with the collection of data from both lower (contralateral) limbs, this initial stage of organization extends to an individual level, encapsulating the unique characteristics of each participant's gait. This aspect holds particular relevance in the context of neurological conditions (Rodríguez-Martín et al., 2017; Ingelse et al., 2022). Subsequently, the aggregation of individual-level datasets contribute to the formation of distinct groups (HC and MS in the case of this work). Finally, group categorizations, in turn, become nested within the broader population of individuals, representing a diverse spectrum of human gait patterns. This approach can offer a comprehensive understanding that considers individual variations, group dynamics, and shared trends across the entire population. However, to the best of the authors' knowledge, the methods currently adopted in the gait-analysis community do not necessarily capture the hierarchical structure of the acquired data. Where many methods presented in the literature will look at a set of summary features—usually computed from data which has been averaged across multiple gait cycles—this article proposes to model the functional form of shank angular velocity across the gait cycle. This approach can be interpreted as a *nonparametric* representation of the gait. The equivalent feature space is expected to be much richer than a low-dimensional one and has the advantage of not requiring "expert" selection of those features.

In view of the gait characteristics of PwMS and considering the distinctive features observed in the dataset depicted in Figure 1, this study pioneers a novel methodology for constructing a robust probabilistic model specifically tailored to the shank angular velocity. The probabilistic approach proposed in this work may potentially provide valuable insights in the field of gait analysis, especially in the challenging context of assessing and quantifying the degree of gait impairment and its changes over time. In the context of neurological disorders, such as MS, which is marked by intrinsically unpredictable disease progression (Gelfand, 2014; Creagh et al., 2022), the proposed probabilistic framework becomes

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

particularly relevant. A probabilistic framework will provide distributions over the expected gait patterns along with a measure of confidence, allowing informed decision-making in data-driven assessments of pathological gait. This is the opposite of a deterministic approach, where uncertainty is not accounted for, and therefore implies perfect models. A deterministic approach can lead to shortcomings in planning for unforeseen variations or complexities in the individual's gait. Furthermore, the hierarchical extension can advance the analysis capabilities, allowing for a granular analysis of the gait patterns. Here, idiosyncrasies of an individual's gait patterns can be immediately captured, together with the corresponding confidence bounds. Moreover, group-level differences can be revealed, as well as isolated. It is therefore clear that a hierarchical probabilistic modeling approach offers tangible benefits for the gait analysis community.

This article aims to provide a methodology for accurately modeling the shank angular velocity through an extension of the hierarchical Gaussian processes (GPs) model proposed by Hensman et al. (2013), which effectively manages heteroscedasticity and facilitates sparse inference. The authors hypothesize that such a model would be able to showcase similar trends consistent with the existing literature on lower limb joint kinematics. Importantly, the model is anticipated to achieve this alignment in a manner that reflects the inherent organization of the dataset. The contribution of this paper can be summarized into three key modeling ideas. First, the hierarchical structure inherent in the data is leveraged to capture the *temporally structured covariance* between contralateral limbs, individual subjects, and groups. This hierarchy accommodates the shared underlying population patterns across both groups, while also accommodating the characteristic group patterns present in all individuals in each group. Additionally, it considers the extension of distinctive individual patterns to contralateral limbs, in order to deal with the potential presence of lower limb asymmetry characterizing the MS-affected group (Pau et al., 2021). Second, given the substantial amount of data collected during clinical assessments, this work addresses scalability challenges through *variational sparse approximations* (Titsias, 2009). This ensures the efficiency of the GP in handling large datasets. Third, recognizing the nonconstant variability of the shank angular velocity across the gait cycle, *heteroscedasticity* is introduced into the GP framework by modeling the variance as an input-dependent function (Lázaro-Gredilla and Titsias, 2011). This approach leads to a sensitive and informative method for characterizing the shank angular velocity patterns.

The article is organized as follows. First, the relevant literature is presented in Section 1.1. Then, the key concepts necessary for the formulation of the hierarchical GP model are presented in Section 2. Following this, the relevant extensions for handling heteroscedasticity and sparse inference are detailed in Sections 2.1, 2.2, and 2.3. Section 3 introduces an example dataset, which will then be utilized to showcase the potential utility of this newly proposed modeling approach in the context of gait analysis. The applications of the methodology on the chosen case study are presented and discussed in detail in Sections 3.2, 3.3, and 3.4. Finally, the article concludes in Section 4.

1.1. Related work

Within the gait analysis community, various approaches have been used for characterizing the gait patterns, broadly falling into three different categories: conceptual gait models, model-based optimization techniques, and statistical methods. Conceptual gait models consist of a finite set of features extracted from the gait signals and are typically initiated with the accurate identification of gait events within the gait cycle. For example, conceptual gait models have been derived for characterizing the gait patterns of people with Parkinson's disease (Arcolin et al., 2019), as well as community-dwelling older adults (Verghese et al., 2009; Lord et al., 2013), people with dementia (Verghese et al., 2007), or MS (Angelini et al., 2021). Specific to the MS population, Shema-Shiratzky et al. (2019) investigated the deterioration of specific aspects of the gait during a typical 6-minute walking test (6MWT), revealing that key metrics —including cadence, stride time variability, stride regularity, step regularity and gait complexity— significantly deteriorated during the test, relative to HCs. Later, Angelini et al. (2020) studied gait alterations for patients with moderate and severe MS. The study also incorporated supplementary gait metrics, such as intensity, jerk, regularity, and symmetry, which provide additional insights into overall gait quality and efficiency (Pasciuto et al., 2017). In a subsequent study, Angelini et al. (2021) proposed

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

the previously mentioned MS-specific conceptual gait model, which comprises metrics categorized into five domains (rhythm/variability, pace, asymmetry, and forward and lateral dynamic balance). This model successfully identified gait abnormalities across different MS disability groups. Nevertheless, a notable challenge in these studies lies in assessing the model's capacity to generalize across unseen datasets. Moreover, by solely focusing on a limited set of features, the authors argue that a vast amount of costly clinical data—potentially encoding important information about the health condition of specific individuals—is discarded.

Model-based optimization techniques have also been used to gain a better understanding of human walking patterns. These approaches leverage idealized mathematical models and cost functions in order to determine an optimized trajectory of the moving body part (Anderson and Pandy, 2001; Neptune et al., 2009; Xiang et al., 2010; Rasouli and Reed, 2021). The optimization problem is formulated as a minimization task over various human performance measures, such as dynamic effort, mechanical energy, metabolic energy, jerk, and stability, while adhering to physical constraints (Xiang et al., 2010). Yet, although informed modeling strategies are utilized, optimization-based techniques can be an oversimplification of the actual gait dynamics, and some are not representative for pathological populations (Rasouli and Reed, 2021). However, the biggest limitation of these approaches is that most optimization techniques output a deterministic pattern, which does not account for the uncertainty associated to the target process (Yun et al., 2014), even though it is well known that human gait involves randomness, which cannot be fully captured (Hausdorff et al., 1995).

Statistical methods offer a viable alternative for predicting gait patterns without the need for predefining biomechanical models. By inherently accommodating variations and uncertainties inherent in human walking, these methods present a promising approach. In light of the diverse range of available gait analysis acquisition systems (including motion capture devices, force plates, and inertial measurement units), a multitude of machine-learning techniques such as neural networks (Bishop, 2006), support vector machines (SVMs) (Cortes and Vapnik, 1995; Smola and Schölkopf, 2004), or Gaussian process regression (GPR) (Rasmussen and Williams, 2006) have established themselves as prominent data-driven tools for effectively handling extensive gait data processing and inference. For example, Horst et al. (2016) used SVMs to monitor daily kinematic changes in individual gait patterns. Later, Horst et al. (2019) used both SVMs and neural networks to investigate specific gait features in the gait patterns that can accurately identify specific individuals. Gadaleta et al. (2016) used a convolutional neural network (CNN) to extract gait features from a single wearable sensor placed on the shank. Later, these features were fed to an SVM classifier for human gait-based person recognition. In a later work, Gadaleta and Rossi (2018) used the same approach to form an outlier detection problem for human gait identification. More recently, Fang et al. (2020) developed a gait recognition and prediction model based on a temporal CNN for improving the interactions between exoskeletons used for rehabilitation purposes and their users. Nevertheless, these approaches do not inherently offer uncertainty estimates for predictions and, as such, a distinction should be made between deterministic and probabilistic approaches.

Within this context, GPs have emerged as a powerful probabilistic tool for gait pattern prediction. GPs are nonparametric models that can capture complex patterns and dependencies, without assuming any specific functional form. Moreover, GPR allows comparisons of the gait cycles across the full *function space* rather than simply on a discrete feature level. Wang et al. (2008) proposed a dynamical GP model for human motion, which was later used by Chun et al. (2015) and Hong et al. (2019) in order to generate reference trajectories for robotic gait rehabilitation systems. Yun et al. (2014) also used GPR for mapping body parameters to gait kinematics. Glackin et al. (2014) attempted to model the lower limb joint kinematics of individual subjects and showed that GPs can learn a mapping between patient's gait and therapist-assisted gait. However, limited conclusions can be drawn from this study, as a result of the limited number of subjects included. Wu et al. (2018) introduced GP regression for learning the relationship between body parameters and gait features at different walking speeds, as part of a developmental pipeline designed for individualized lower limb exoskeleton robots, while Chen et al., 2023 used deep GPs for online gait prediction during human–exoskeleton interaction. In a different context, Benemerito et al. (2022) introduced GPs as a regression tool for efficient sensitivity analysis

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

aimed at reducing the complexity of musculoskeletal modeling, in response to the computational challenges offered by traditional Monte Carlo methods. It can be seen that there have been numerous approaches toward modeling gait patterns and that investigation into this problem is an active field of research. While most of the studies have been directed toward robotics applications utilizing modeling of healthy patterns, GPR adoption for modeling pathological gait patterns remains limited. Additionally, to the best of the author's knowledge, no studies have specifically attempted the probabilistic modeling of shank angular velocity. Moreover, the challenges associated with the scalability of GPs are seldom addressed (Chen et al., 2023), and the integration of heteroscedastic approaches remains largely unexplored for this specific application. Furthermore, it is also significant to highlight that hierarchical extensions that mimic the organization of the data have not been ventured within the realm of gait analysis, as far as the authors are aware.

To address these challenges, motivated by the characteristics of the example dataset presented in Section 1, this paper provides a novel methodology that forms an accurate hierarchical model for the shank angular velocity kinematics. Concurrently, the varying uncertainty is automatically quantified across the duration of the gait cycle in a manner that is efficient for large datasets and statistically rigorous. The approach proposed in this article integrates the methodologies of Liu et al. (2018), which enables sparse inference along with handling of heteroscedastic noise, with a novel application of hierarchical modeling, as originally proposed by Hensman et al. (2013).

2. Gaussian process regression

Gaussian process (GP) models represent a versatile nonparametric Bayesian machine learning approach for resolving regression problems, enabling the characterization of a distribution over functions (Rasmussen and Williams, 2006). Specifically, a GP is an infinite set of random variables that exhibit a joint Gaussian distribution for any finite subset. GPs have gained significant popularity across a diverse range of applications owing to their ability to automatically quantify uncertainty in predictions, minimal requirement for a priori input, and modeling capabilities even in the presence of high noise levels in the measured data. The GP is developed to model data as the output of some function $f(\boldsymbol{x})$, operating on a D-dimensional input $\boldsymbol{x}$, as described by Equation 1.

$$y = f(\boldsymbol{x}) + \varepsilon, \varepsilon \sim \mathcal{N}\left(0, \sigma_n^2\right) \tag{1}$$

Here, it is assumed that the measured values $\boldsymbol{y}$ differ from the latent function values $f(\boldsymbol{x})$ by some additive noise ε with zero mean and a predetermined variance σ_n^2, also known as the "nugget parameter." Equation 2 formally defines a GP, where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are a pair of inputs to the function of interest. For clarification, the notation adopted in this study involves representing vectors using bold typography, while matrices are identified by uppercase letters.

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')) \tag{2}$$

It follows that a GP is completely specified by its mean function $m(\boldsymbol{x})$ and the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$. Here, the covariance function encodes the similarity between any pair of inputs. A popular choice for the covariance function is the $3/2$ Matérn kernel, which is described in Equation 3. Although other kernel functions exist, such as the squared exponential kernel, the Matérn kernel was selected, as it was found to better model the abrupt changes in the slope of the gait traces, as a result of its finitely differentiable property.

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \left(1 + \frac{\sqrt{3}(\boldsymbol{x} - \boldsymbol{x}')^2}{l}\right) \exp\left(\frac{-\sqrt{3}(\boldsymbol{x} - \boldsymbol{x}')^2}{l}\right) \tag{3}$$

Note that this kernel function is defined by a set of two hyperparameters: the variance σ_f, controlling the vertical scaling (amplitude) of the kernel and the length-scale l, which controls the smoothness of the functions. Next, the prediction task is achieved by assessing the joint Gaussian distribution of the observed target values and the function values at the test locations.

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

Finally, in order to learn the hyperparameters, a Type-II maximum likelihood (ML-II) approach is used by maximizing the *marginal likelihood* of the model. However, for convenience and numerical stability, the optimization is performed as a minimization task over the *negative log marginal likelihood.* For the specific mathematical details regarding GP implementation, the reader is referred to Rasmussen and Williams (2006) and to Appendix A, where the key equations are presented.

2.1. Sparse GPs for large datasets scalability

Either learning the hyperparameters of the GP or making predictions involves taking the inverse of the covariance matrix with noise, $\left(K_{xx}+\sigma_n^2\mathbb{I}\right)^{-1}$, which is an operation scaling as $\mathcal{O}\left(N^3\right)$ in computational complexity. Hence, in practice, it is not feasible to perform full GP regression tasks on datasets involving more than roughly 10,000 datapoints (Rogers et al., 2020). This is also one of the limitations preventing the use of full GP regression on gait data, as the number of datapoints collected during a visit often exceeds 10,000 points per subject. To address this limitation, a number of approximation methods have already been proposed in the literature (Quiñonero-Candela and Rasmussen, 2005; Titsias, 2009; Bui et al., 2017; Hensman et al., 2018). Broadly speaking, these approaches are divided into two main classes, namely *model approximations* and *posterior approximations.* For the sake of brevity, the reader is referred to Quiñonero-Candela and Rasmussen (2005), Rasmussen and Williams (2006), or Bui et al. (2017) for more details about these approaches. The posterior approximation approach is widely recognized to generally provide more robust approximations and possess inherent mechanisms to counteract overfitting. Thus, the present study employs a posterior approximation method, specifically the variational free energy (VFE) method proposed by Titsias (2009). The main advantage of this approximation method is the reduction in time complexity from $\mathcal{O}\left(N^3\right)$ to $\mathcal{O}\left(NM^2\right)$, where M is the number of auxiliary points introduced, called *inducing points*, at which the approximation is performed. Clearly, this becomes advantageous when $M \ll N$. Therefore, with this approximation method, the standard GP can be scaled up to large datasets, such as those containing gait data collected during clinical assessments for multiple patients.

The *variational approximation* of the full posterior is handled through the use of a small set of inducing points, $\{Z, \boldsymbol{u}\}$ (where Z contains the locations of the inducing points and $\boldsymbol{u}$ are the values of the latent functions at these points). The model can then be learnt by minimizing the Kullback–Leibler (KL) divergence between the approximate joint posterior and the full joint GP posterior. Nonetheless, this minimization is equivalent to maximizing a variational lower bound (also known as the *Evidence Lower Bound* or *ELBO*) of the true log marginal likelihood, as detailed in Titsias (2009). For conciseness, the key equations characterizing the variational approximation method used in this paper as a means of scaling GPs to large datasets can be found in Appendix B.

2.2. Sparse heteroscedastic noise models extensions

With reference to the dataset presented in Figure 1, it is evident that the *homoscedastic* noise assumption of the GP model is not satisfied. This is the assumption that the additive noise on the function $f(\boldsymbol{x})$ has a constant variance across the input space. To address this problem, *heteroscedastic* GP models (that is those using input-dependent additive noise) have been developed (Lázaro-Gredilla and Titsias, 2011). Focusing on the MS group, it can be seen that gait pattern variability is one of the key aspects of the disease. Hence, enhancing the model's ability to effectively capture the inherent variability, particularly in the swing phase of the gait cycle, will enable the establishment of more accurate confidence bounds for predictions. In this case, the regression model introduced by Equation 1 would then become

$$y = f(\boldsymbol{x}) + \epsilon(\boldsymbol{x}), \epsilon(\boldsymbol{x}) \sim \mathcal{N}(0, r(\boldsymbol{x})) \tag{4}$$

This means that the variance of the noise process is now a function of the model inputs. Notably, the *heteroscedastic* noise model presented above reduces to a *homoscedastic* one when $r(\boldsymbol{x})$ is a constant. The derivation of the heteroscedastic GP model was first introduced by Lázaro-Gredilla and Titsias (2011). To

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

define this model, a GP prior is initially placed on the unknown function $f(\boldsymbol{x})$, as presented in Equation 5. Here, a zero-mean function has been assumed for practicality reasons, although this assumption is not restrictive, and alternative mean functions can be considered.

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(0, k_f(\boldsymbol{x}, \boldsymbol{x}')\right) \tag{5}$$

Then, to ensure positivity of the noise variance, $r(\boldsymbol{x})$, an exponential transform is applied, as described in Equation 6:

$$r(\boldsymbol{x}) = \exp(h(\boldsymbol{x})), \quad h(\boldsymbol{x}) \sim \mathcal{GP}(\mu_0, k_h(\boldsymbol{x}, \boldsymbol{x}')) \tag{6}$$

Here, the log noise variance, denoted by $h(\boldsymbol{x})$ is modeled by a GP, whose covariance function is denoted by $k_h(\boldsymbol{x}, \boldsymbol{x}')$. The GP has a constant mean, μ_0, which controls the *scale* of the noise process. Here μ_0 is introduced as a learnable parameter, which models the "average" noise level across the function. This is a departure from standard function modeling where $\mu_0 = 0$ is typically assumed. By learning μ_0, we can account for the inherent noise present in the data, which corresponds to the homoscedastic case. As such, the addition of the secondary GP placed on the log noise variance increases the expressiveness of the GP, albeit simultaneously increasing the complexity of the learning and inference processes. As a result of the inclusion of the heteroscedastic noise model, the log-likelihood becomes untractable and can no longer be computed analytically. Similarly to the sparse GP extension, a variational method is used to approximate the posterior distribution and form a new lower bound. The key equations describing the lower bound and the variational approximation of the first two moments of the predictive distribution are detailed in Appendix C.

The concepts outlined above facilitate the probabilistic estimation of latent underlying functions, along with accurate prediction of variance, at any given input. However, in practice, computing the variational bound of the heteroscedastic GP, along with its gradient, takes roughly twice the time required to compute the evidence and its derivatives in a standard homoscedastic GP (Lázaro-Gredilla and Titsias, 2011). Due to this cost consideration, integrating a variational sparse heteroscedastic method to improve scalability becomes a necessity. To this end, the Variational Sparse Heteroscedastic Gaussian Process (VSHGP) has already been derived in the literature by Liu et al. (2018). Inspired by the model presented in Lázaro-Gredilla and Titsias (2011), Liu et al. (2018) have shown that it is possible to derive an analytical *ELBO* using M inducing points for the mean function GP, $f(\boldsymbol{x})$, and U inducing points for approximating the log noise variance GP modeling $h(\boldsymbol{x})$. As a result, this approach is scalable to large datasets, given its $\mathcal{O}\left(NM^2 + NU^2\right)$ complexity.

Finally, with these methods it is now possible to approximate the first two moments of the predictive distribution, that is the mean and variance, respectively. The nontrivial key equations leading to the derivation of the first two moments can be found in Appendix D. It is also worth noting that, in addition to the methodology presented here, Liu et al. (2018) also improved the scalability of the proposed VSHGP model through the addition of stochastic and distributed extensions. However, these extensions are not utilized in this study. For more details regarding these, the reader is instead referred to the original paper. As it will later be demonstrated, the additional information provided by the VSHGP model regarding the uncertainty of the process will prove to be an important capability for modeling kinematic gait patterns, especially in pathological populations.

2.3. Hierarchical expansion

In this study, we reassess the dataset depicted in Figure 1, comprising two distinct subgroups: HCs and PwMS. The final modeling strategy employed herein involves a hierarchical extension of the VSHGP model presented in Section 2.2. This extension entails its integration with the hierarchical model proposed by Hensman et al. (2013). Specific to the current dataset, the key idea of hierarchical models is that there exists a common trend across a pool of data from multiple candidates performing the same walking test, regardless of their group label. The measurements obtained from each participant exhibit individual

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

variation from the shared trend due to biomechanical differences, as well as corruptive noise. Consequently, these distinctive individual gait patterns can be interpreted as individual signatures (Winner et al., 2023). For the sake of simplicity, this section will only consider a two-layer hierarchy and will only present the implementation using a standard GP. The integration of the variational sparse heteroscedastic extension in a hierarchical framework will be explained further in this section of the report by considering block-wise relationships within the hierarchical covariance.

Let $\boldsymbol{y}_{gi}$ denote the vector of measurements for the ith individual in group g. The corresponding time points are stored in vector $\boldsymbol{x}_{gi}$. To combine the data acquired during all the walking tests from all participants in a particular group, a Bayesian hierarchical approach is being used. The underlying trend for the gth group, $f_g(\boldsymbol{x})$ is presumed to be drawn from a GP with a zero-mean and whose covariance function is denoted by $k_g(\boldsymbol{x},\boldsymbol{x}')$. Further down in the hierarchy, the underlying trend describing the gait pattern belonging to a unique participant in that particular group, $f_{gi}(\boldsymbol{x})$ is drawn from the group GP. However, the mean of the individual GP is $f_g(\boldsymbol{x})$, as described in Equation 7.

$$\begin{gathered} f_g(\boldsymbol{x}) \sim \mathcal{GP}\left(0, k_g(\boldsymbol{x},\boldsymbol{x}')\right) \\ f_{g,i}(\boldsymbol{x}) \sim \mathcal{GP}\left(f_g(\boldsymbol{x}), k_i(\boldsymbol{x},\boldsymbol{x}')\right) \\ \text{where } i = 1,2,\ldots,n \end{gathered} \tag{7}$$

It should be noted that the two covariance functions k_g and k_i used for the group and individual levels may be different. For clarity, while group kernel hyperparameters remain constant across all individuals, individual kernel hyperparameters are allowed to vary, reflecting specific individual variability. This model is pictorially shown in Figure 2, where the function dependency is highlighted. The prior over the underlying group pattern $f_g(\boldsymbol{x})$ is shown at the top, as a dotted line. The shaded area represents the variance of the function, which is controlled by σ_g^2. The smoothness of the function is controlled by the length-scale of the group kernel, l_g. A single sample from this prior is then shown as a red solid line,and the length-scale of the covariance function is also highlighted. The individual level is shown in the second row, where samples conditioned on the sample shown in $f_g(\boldsymbol{x})$ are displayed, representing three unique individuals. The three samples follow the trend of $f_g(\boldsymbol{x})$, but are allowed to independently vary by a small amount (σ_i^2) with a short length-scale l_i. Therefore, although the main features of the common trend are preserved, each of the individuals exhibit their own characteristics. Finally, the hierarchical covariance matrix is shown at the bottom of Figure 2, demonstrating the block-wise relationship between individuals.

Let $\boldsymbol{Y_g} = \left\{\boldsymbol{y}_{gi}\right\}_{i=1}^{n}$ be the collection of noisy observations for n patients in group g and $\boldsymbol{X_g} = \left\{\boldsymbol{x}_{gi}\right\}_{i=1}^{n}$ the corresponding time points. Due to the conjugacy property of Gaussian distributions, the model described above can be mathematically represented as a joint Gaussian distribution, and it is possible to write down the likelihood as

$$p\left(\boldsymbol{Y_g}|\boldsymbol{X_g},\theta\right) \sim \mathcal{N}\left(\widehat{\boldsymbol{y}}_{\boldsymbol{g}}|\boldsymbol{0},\Sigma_g\right) \tag{8}$$

where $\widehat{\boldsymbol{y}}_{\boldsymbol{g}} = \left[\boldsymbol{y}_{g,1}^T, \boldsymbol{y}_{g,2}^T, \ldots, \boldsymbol{y}_{g,n}^T\right]^T$ has been used to denote the row-wise concatenation of $\boldsymbol{Y_g}$. θ are the hyperparameters of the covariance functions $k_g(\cdot)$ and $k_i(\cdot)$. Finally, the block of Σ_g is given by

$$\Sigma_g[i,i'] = \begin{cases} K_g\left(\boldsymbol{x}_{gi},\boldsymbol{x}_{gi'}\right) + K_i\left(\boldsymbol{x}_{gi},\boldsymbol{x}_{gi'}\right) + \sigma_n^2\mathbb{I} & \text{if } i = i' \\ K_g\left(\boldsymbol{x}_{gi},\boldsymbol{x}_{gi'}\right) & \text{otherwise.} \end{cases} \tag{9}$$

In order to make inferences about the functions $f_g(\boldsymbol{x})$ and $f_{g,i}(\boldsymbol{x})$, it is necessary to compute the covariances between these functions. The predictive covariance functions are given in Equation 10. Note that these expressions describe the covariance of the latent functions, while the covariance of the observed data y_{gi} additionally includes the heteroscedastic noise variance $r(x)$ via moment-matching (Lázaro-Gredilla and Titsias, 2011). This means that group predictions can be made simply by using the group kernel, $k_g(\cdot)$, whereas an additive kernel, $k_g(\cdot) + k_i(\cdot)$, is required for individual predictions.

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

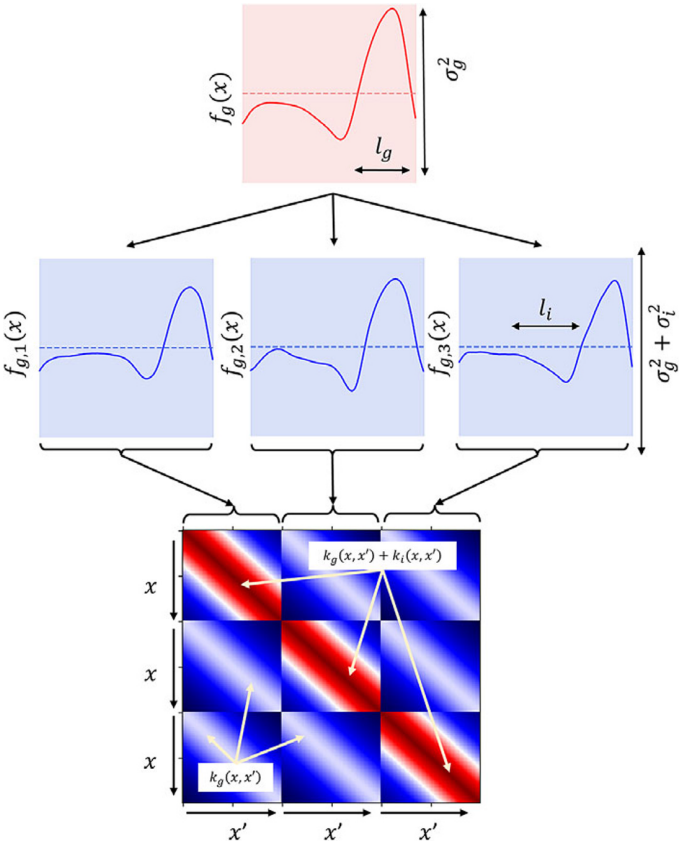

Figure 2. *An illustration of a simple hierarchical GP. Top: solid line—a single sample from the prior over the underlying group function $f_g(x)$. Dotted line—zero-mean function. Shaded area: ± 1 standard deviation of functions, σ_g^2. Middle: three samples conditioned on $f_g(x)$ and corresponding to three distinct individuals. The individual samples follow the trend of $f_g(x)$, but vary by a small amount, σ_i^2. The length-scale of the group and individual functions are denoted by l_g and l_i, respectively. Bottom: block-wise covariance matrix used to generate samples.*

$$cov\left(f_{gi}(x), f_g(x')\right) = k_g(\boldsymbol{x},\boldsymbol{x}'), \tag{10a}$$

$$cov\left(f_{gi}(x), f_{gi'}(x')\right) = \begin{cases} k_g(\boldsymbol{x},\boldsymbol{x}') + k_i(\boldsymbol{x},\boldsymbol{x}') & \text{if } i = i' \\ k_g(\boldsymbol{x},\boldsymbol{x}') & \text{otherwise.} \end{cases} \tag{10b}$$

Finally, inference can then be made using the standard methods outlined in the preceding sections, while hyperparameters of the covariance functions can also be optimized in a similar fashion. However, the scalability of the hierarchical model is similar to that of a standard GP, limiting its applicability to large datasets. Hence, to mitigate the computational challenges associated with large datasets, variational approximation methods have been employed to efficiently approximate the posterior distributions, as described in Sections 2.1 and 2.2. The hierarchical formulation employs a shared set of inducing points across all GPs in the hierarchy, in contrast to maintaining distinct sets for each hierarchical level. This formulation serves to both reduce computational complexity and constrain the parameter space, as it requires optimizing a single shared set of inducing point locations rather than multiple disjoint sets across the hierarchy.

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

3. A case study for characterizing gait variability in MS using hierarchical Gaussian processes

Here, it remains to demonstrate how the above proposed scheme might be applied to gait data collected from both patients and healthy control individuals. The purpose of this case study is to highlight how the framework for modeling of gait (with heteroscedastic noise) provides insights from all levels of the hierarchy, from specific limbs to populations. The examples shown in this section will progress downward from the top of the hierarchy, starting with comparisons between groups of PwMS and HCs, and finishing with considerations of symmetry in individuals' gait. The dataset used in this work consists of angular velocity data recorded with IMUs from 2 groups of subjects: namely healthy controls (HCs, with no history of musculoskeletal or neurologic disorders which might affect their balance or mobility) and individuals with MS. Both groups were comprised of 28 subjects each. The severity of MS was assessed through the Expanded Disablity Status Scale (EDSS) scale (Kurtzke, 1983), which is one of the most widely used clinical outcomes (Angelini et al., 2021), consisting of a neurological assessment, as well as observing the walking range and the level of walking assistance needed. The scale is rated from 0 (normal healthy status) to 10 (MS-related death) in 0.5-unit increments. Scores up to 3.5 typically indicate no visible gait impairment, while scores between 4.0 and 5.5 denote individuals capable of walking limited distances independently. Scores up to 6.5 indicate the necessity of assistive walking devices, while higher scores denote restricted mobility. In this case study, participants with relapse-remitting MS were included only if they had experienced no relapse in the 30 days preceding the baseline test and had maintained a stable treatment for the past 3 months. No participants using assistive devices were included. Ethics approval was granted by both the NRES Committee Yorkshire & The Humber-Bradford Leeds (Ref: 15/YH/0300) and the North of Scotland Research Ethics Committee (Ref: 17/NS/0020). Written informed consent was obtained from all participants prior to their inclusion in the study.

Gait data was acquired using two tri-axial IMUs (OPAL, APDM Inc, Portland, OR, USA, sampling frequency, 128 Hz, gyroscope range $\pm 2000\,^\circ/\mathrm{s}$), attached to the body through elastic straps, on the anterior aspect of both shanks. The sensing axes of the sensors were approximately aligned with the anatomical planes. Both groups (HC and MS) performed the 6MWT, going back and forth in a straight line across a corridor of either 10 or 14 m and turning at the ends (Table 1). Participants were instructed to walk at their self-selected pace, with rests permitted only if necessary. All turns and resting breaks were automatically removed, following the procedure detailed in Angelini et al. (2021). Only straight-line walking bouts were included in the upcoming analysis. Data was segmented into individual strides, according to the gait events locations identified in the shank angular velocity signal (Angelini et al., 2021). To remove sensor misalignment effects, a rotation to a vertical-horizontal coordinate system was applied, as described in Moe-Nilssen (1998). The shank angular velocity signals were filtered with a zero-phase, low-pass, Butterworth filter with a 10 Hz cut-off frequency and normalized using a zero-mean, unit range normalization method. It was decided to remove the transient part at the beginning of the signals (first 8% of the samples), in order to avoid problems caused by misclassification of the gait events (Haji Ghassemi et al., 2018). Finally, for each individual limb, the resultant gait cycles were normalized along the time axis, effectively eliminating the pace component from the signals and facilitating direct comparative analysis.

Table 1. *Demographics table*

		Age	Gender	MS subtypes	
Group	Participants	Mean (SD)	*N* male	RR	SP
HC	$n = 28$	39.2 (12.7)	13	–	–
MS	$n = 28$ $\overline{EDSS} = 3.35$	47.7 (12.2)	8	18	10

Note. RR, relapse remitting; SP, secondary progressive.

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

3.1. Case study objectives and model evaluation metrics

Given the anticipated differences among the gait patterns of HCs and PwMS, the aim of this case study is to showcase the intrinsic advantages of the proposed modeling approach. Therefore, having established a mathematical framework for hierarchically modeling the nonlinear gait pattern, together with heteroscedastic noise, the focus can now shift toward predicting the shank angular velocity for the two groups considered in this study. The following results will serve to highlight the benefits of the three main aspects present in the proposed novel modeling methodology. First, it is necessary to establish that the uncertainty found in the gait data is heteroscedastic, warranting this more complex modeling choice and demonstrating the insight that can be learnt from the functional form of the noise variance. Second, it will be shown how the hierarchical nature of the model naturally obeys the structure arising in gait data when aggregating different limbs, individuals, and groups (e.g. MS and HC), and facilitates quantitative comparisons across multiple scales. Finally, considering that angular velocity data was acquired from both limbs, as well as the time normalization procedure used in this study, the third objective of this study is to showcase the utility of the probabilistic models by introducing a novel methodology for lower limb asymmetry quantification. With regards to the third aim of this study, it is important to note that the concept of asymmetry presented here deviates from the traditional concept of gait asymmetry, which is commonly assessed as the absolute difference in temporal metrics between contralateral limbs or as the natural logarithm of the absolute ratio between their mean values (Godfrey et al., 2015; Yogev et al., 2007). However, the full motivation for alternative asymmetry metrics is postponed until Section 3.4.

A total number of 50,015 gait cycles has been collected, (26,330 belonging to healthy individuals and 23,685 to PwMS). Following pre-processing, the data are separated into two distinct sets: the *training* set and the *held-out* test set. The training set consists of 70% randomly selected samples from the aggregate dataset, that is combined data from all subjects. For clarity, a single sample refers to a single data point. Only this set was used for *training* and hence the optimization of the hyperparameters. The test set contains the remaining 30% of the data, that is these samples remain *unseen* until predictions are made and are referred to as the *held-out* test set. The training set contains 1,942,881 data points, while the test set contains 832,664 points. It can be therefore seen that given the very large datasets, it is not feasible to fit non-variational sparse GPs with any reasonable amount of computational resources. In addition to this, training the GP models on a subset of data (that is downsampled data) may lead to unreliable uncertainty quantification, which fails to capture the full information present in the complete dataset (Quiñonero-Candela and Rasmussen, 2005). This is particularly important in gait analysis, where both the overall pattern and subtle variations can be clinically relevant. As such, for full transparency, comparisons with standard GP formulations have not been conducted in this study.

The hierarchical variational sparse heteroscedastic GP (HVSHGP) model, combining the hierarchical approach, together with sparsity and heteroscedasticity was implemented in Python, using GPflow (de G. Matthews et al., 2017). As an implementation note, instead of constructing the block-wise covariance matrix, the underlying dependencies and interdependencies between different subjects and groups were modeled by incorporating a block-wise relationship into an additive *ELBO*, which was then optimized using the NADAM (Nesterov-accelerated Adaptive Moment Estimation) algorithm (Dozat, 2016)—an extension of the Adam optimizer that combines Nesterov momentum with adaptive moment estimation to improve convergence speed and stability. Optimization was performed with a learning rate of 0.0001 for 1000 steps, and default parameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$. While the derivatives of the lower bound (see Equation 30) with respect to the hyperparameters are provided in the framework proposed by Liu et al. (2018), the additive ELBO used in this hierarchical approach was optimized using the automatic differentiation capabilities of TensorFlow (Abadi et al., 2015). Moreover, it should be noted that optimizing over the size of the inducing point set represents a challenging high-dimensional optimization problem. As such, while the size of the inducing point set is fixed at 100 across all levels of the hierarchy, their locations are treated as optimization variables. This represents a trade-off between computational complexity and model fidelity: fixing the number of inducing points helps manage computational cost, while allowing their locations to be optimized enables the model to better capture the underlying structure of the gait patterns. The optimization of inducing point locations does increase the complexity of the variational optimization problem, but we found this additional flexibility to be beneficial for capturing

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

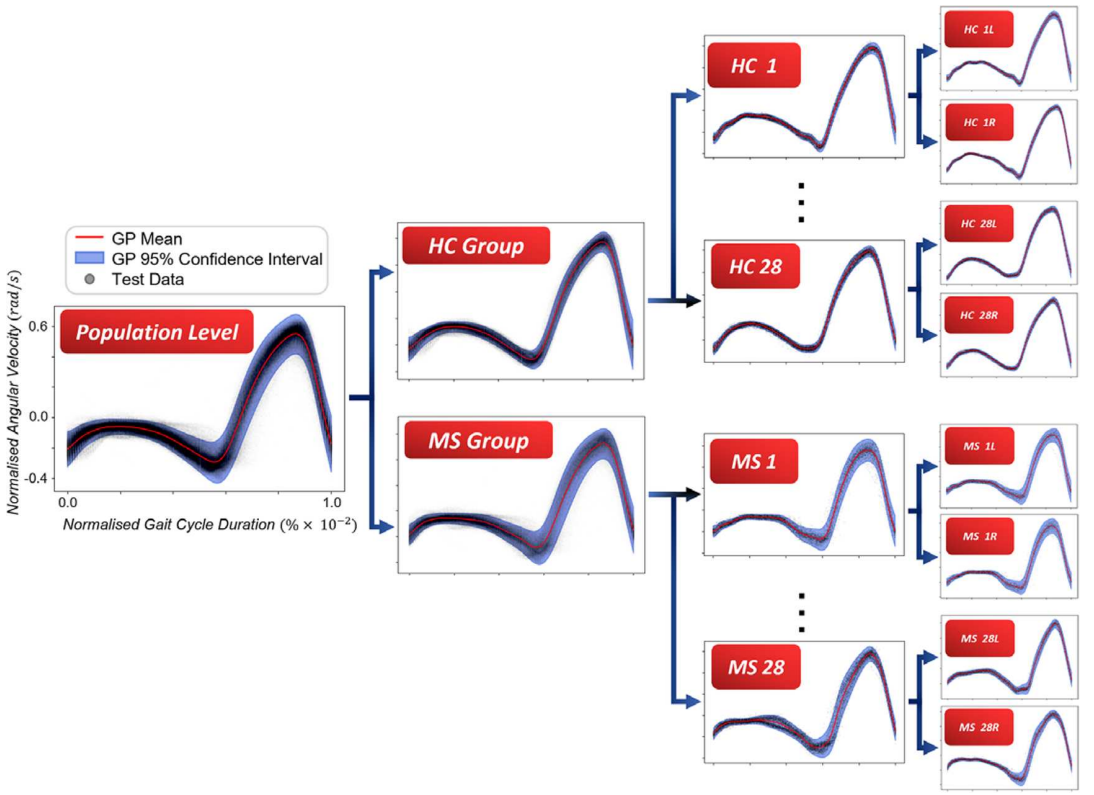

Figure 3. *Hierarchical modeling structure. From left to right: population layer, group layer (HC/MS), individual layer (combining the individual left and right limbs), individual limb layer (modeling the left and right limbs separately).*

nuanced gait patterns. For a more thorough discussion about the selection and optimization of the inducing points set, the readers are referred to Quiñonero-Candela and Rasmussen (2005) or Bui et al. (2017). As a result, a four-layer HVSHGP model was achieved (see Figure 3), and with every layer in the hierarchy, new hyperparameters were introduced. At the top layer of the hierarchy, the entire population is being modeled in order to capture the overall trend shared between both groups. The mean function at the group level is then conditioned on the population samples, and consequently, all the individual mean functions are then subsequently conditioned on their corresponding group samples. Finally, a fourth layer was introduced to manage asymmetric gait patterns corresponding to contralateral limbs.

For the purposes of model comparisons, two performance metrics were used: the *Normalized Mean Squared Error* (NMSE) and the *Mean Standardized Log Loss* (MSLL), which is derived from the *Standardized Log Loss* (SLL), providing a probabilistic measure (Rasmussen and Williams, 2006). The NMSE is computed as

$$NMSE = \frac{100}{n\sigma_y^2}(\boldsymbol{y}_\star - \boldsymbol{y})^T(\boldsymbol{y}_\star - \boldsymbol{y}) \tag{11}$$

where n is the sample size, σ_y^2 is the signal variance, $\boldsymbol{y}_\star$ is the model prediction, and $\boldsymbol{y}$ is the true measured data. Here, an NMSE score of zero implies perfect prediction, meaning that model predictions precisely align with the target values, while a value of 100 corresponds to predicting with the mean of all observations. Therefore, the NMSE measures the average squared difference between the predicted and actual values, normalized by the variance of the target variable, returning the accuracy of the model's predictions and providing an indication about how much the variance in the target variable is captured by

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

the model. Given that GP predictions are returned as distributions, it is feasible to evaluate the negative probability of a prediction under the model, often referred to as the model's loss. By standardizing this value in relation to the mean and variance of the training set, the MSLL can then be computed as

$$MSLL = \frac{1}{n}\sum_{k}\left\{-\log p\left(\boldsymbol{y}_{\star,k}|X,\boldsymbol{y},\boldsymbol{x}_{\star,k}\right) + \log p\left(\boldsymbol{y}_{\star,k};\mathbb{E}(\boldsymbol{y}_k),\mathbb{V}(\boldsymbol{y}_k)\right)\right\} \tag{12}$$

where k indexes a particular test point, $\log p\left(\boldsymbol{y}_{\star,k}|X,\boldsymbol{y},\boldsymbol{x}_{\star,k}\right)$ is the log predictive likelihood of the model, $x_{\star,k}$ represents the test location, X is the set of training inputs, and $\boldsymbol{y}$ are the training targets. Therefore, the MSLL is obtained by taking the average of the negative log likelihood over the test set and subtracting the trivial model, which always predicts the mean and variance of the training set, therefore providing a quantitative metric for how well the model quantifies the uncertainty in predictions. A MSLL value of zero is associated with predicting with the training set mean and variance, and increasingly negative values indicate improved predictions (Rasmussen and Williams, 2006). By using both NMSE and MSLL, users can gain a better, comprehensive understanding of the model's performance.

All statistical comparisons presented in the subsequent sections of the article were performed using GraphPad Prism v.9.5.1 software (GraphPad Inc., La Jolla, CA, USA).

3.2. *Homoscedastic* versus *heteroscedastic modeling for gait data*

In this section, the first objective of the case study will be addressed, which involves showcasing the added benefits of integrating heteroscedastic noise models into GP regression. To support the proposed modeling approach, group-level comparisons will be presented between a non-heteroscedastic four-layer hierarchical variational sparse GP model (HVSGP) and the four-layer HVSHGP model proposed in this work, incorporating the heteroscedastic noise. The model predictions on the test dataset can be seen in Figure 4, for both HC and MS groups, where the first row of predictions belongs to the homoscedastic model, and those on the second row belong to the heteroscedastic model. Qualitatively, it can be seen that the

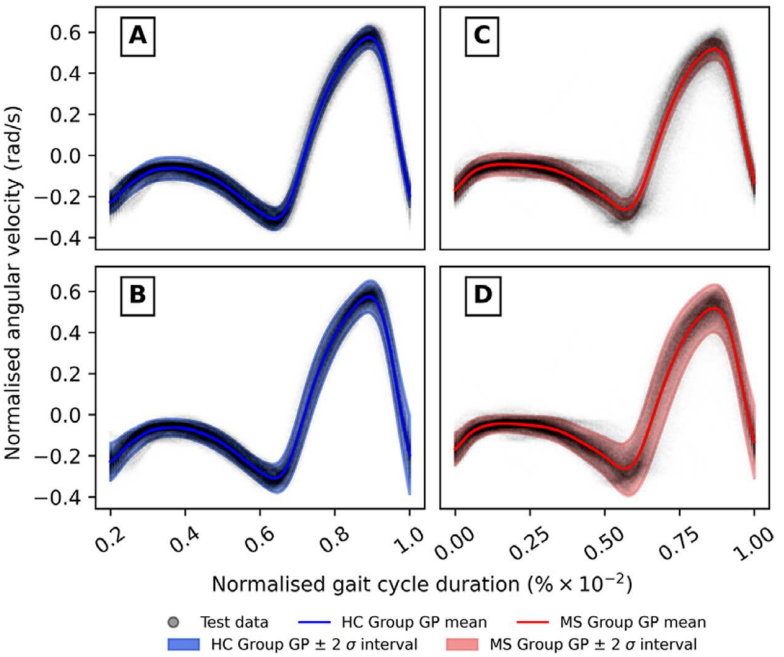

Figure 4. *Group-level GP predictions against test data for homoscedastic models (first row) and heteroscedastic models (second row). (a) HC Homoscedastic model, (b) HC Heteroscedastic model, (c) Homoscedastic MS model, (d) Heteroscedastic MS Model.*

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

homoscedastic model has failed at capturing the representative spread of the data, especially in the case of the MS group, where many test datapoints lie outside the $\pm 2\sigma$ interval. Particularly, the increased variability in the gait signals is not captured adequately during the swing phase of the gait cycle, which accounts for approximately 33 to 38% of the gait cycle (Angelini et al., 2021; Moon et al., 2017; Pau et al., 2017). According to Remelius et al. (2012), this behavior may be a result of unstable equilibrium during the coordination of the swing and forward movement. Conversely, the heteroscedastic model further expands the $\pm 2\sigma$ interval of the GP and allows for better predictions across the entire input space. Even though both models appear to be comparable during the stance phase, the heteroscedastic model also improves the predictions toward the termination of the stance, during the double support time. For clarification, the averaged stance and swing phases for both HC and MS groups can be visualized in Figure 6, which are separated by the second vertical line corresponding to the toe-off event.

The improvement of the heteroscedastic model, relative to the homoscedastic model, is most evident when examining the predictive mean and variance of the models for both HC and MS, as illustrated in Figure 5. Considering only the means, a discernible similarity is observed between the homoscedastic and heteroscedastic approaches, irrespective of the group membership. This behavior was perhaps expected, given that the predictive mean of the heteroscedastic model has a very similar formulation to the predictive mean equation for the homoscedastic case. However, the key distinction between the models, from a predictive standpoint, lies in the computation of the predictive variance, which is learnt via marginal likelihood optimization. This is particularly evident prior to the onset of the swing phase. Notably, an increased variance magnitude is evident for the MS group, surpassing that of the HC group. Consequently, in the MS group, the variability arising from uncontrolled coordination during the swing phase is effectively captured within the $\pm 2\sigma$ bounds of the heteroscedastic GP model. This results in a more accurate representation of group-level gait patterns for this pathology.

The performance metrics of both the homoscedastic and heteroscedastic models are presented in Table 2, concisely quantifying the benefits of the heteroscedastic noise modeling approach. First, it is evident that there are minimal distinctions in model performance between the training and test sets. Next, analyzing the NMSE values, it is unsurprising that both modeling strategies achieved similar results. This is because the NMSE does not depend on the predictive variance of the model. Thus, only looking at this performance metric might be misleading. However, when analyzing the MSLL values, the improved uncertainty quantification of the heteroscedastic model is evident, judging by the negative value of the MSLL. While the addition of heteroscedastic noise modeling led to a modest enhancement in predictive performance for the HC group, the MS group exhibited a notably more significant improvement. This underscores the efficacy of the heteroscedastic approach, particularly for pathological populations where accurate uncertainty estimation is paramount.

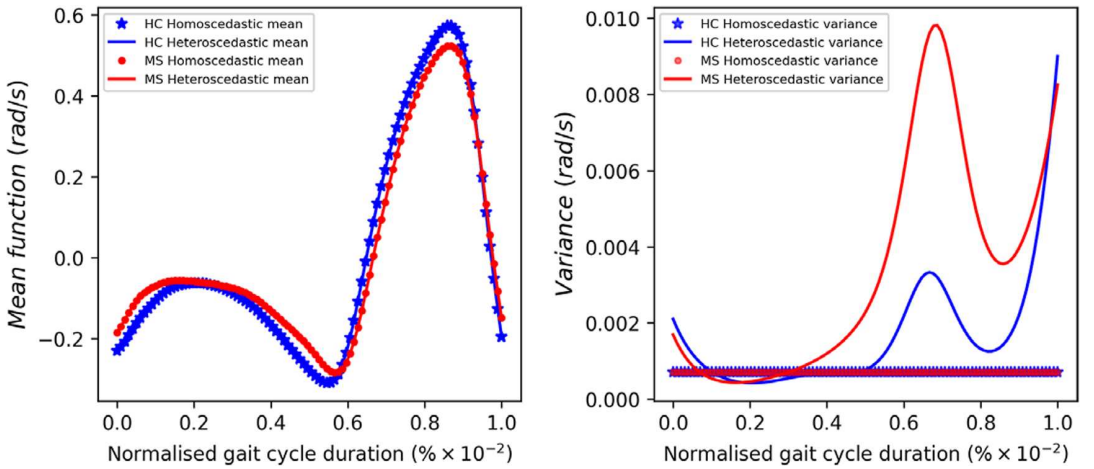

Figure 5. *Mean and variance differences between the homoscedastic and heteroscedastic models.*

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

Table 2. *Comparison of the performance metrics for the homoscedastic and heteroscedastic group models*

	NMSE (%) Train		NMSE (%) Test		MSLL Train		MSLL Test	
	HC	MS	HC	MS	HC	MS	HC	MS
Homoscedastic	2.869	7.403	2.873	7.397	−1.305	0.532	−1.305	0.527
Heteroscedastic	2.869	7.403	2.873	7.397	−1.494	−1.520	−1.493	−1.523

3.3. HC versus *MS: Group and individual level comparisons*

Building upon the utility of heteroscedastic noise modeling, this section targets the second objective of the case study, which aims to present a robust methodological foundation for the systematic examination and quantitative analysis of differences in shank angular velocity patterns across multiple scales. First, the investigation of group-level discrepancies between HC and MS is presented. This investigation serves to establish a comprehensive understanding of the model's capacity to encapsulate the variability inherent in MS gait dynamics. In addition, this comparison serves as a checkpoint, ultimately verifying previously reported literature trends. Expanding on the hierarchical structure of the model, and contrary to conventional approaches that predominantly focus on group-level differences, our exploration extends to facilitate nuanced comparisons at lower levels in the hierarchy, enabling comparisons between individuals.

With reference to the HC and MS group predictions, using the held-out test data, the achieved NMSE scores are 2.873 and 7.397, respectively, indicating a marginal improvement in point-wise error for the HC group. This result may be attributed to the increased measured variance in the MS group, just before the initiation of the swing phase. Notably, the MSLL scores for both group predictions were relatively similar, achieving values of −1.493 for the HC group prediction and −1.523 for the MS group prediction. Next, a more informative analysis may entail a focused examination of specific regions within the gait cycle that manifest the most pronounced differences.

Considering Figure 6, one way of quantifying the differences is by computing the unbiased formulation of the *Maximum Mean Discrepancy* (MMD) (Gretton et al., 2012) at each test location. This is done

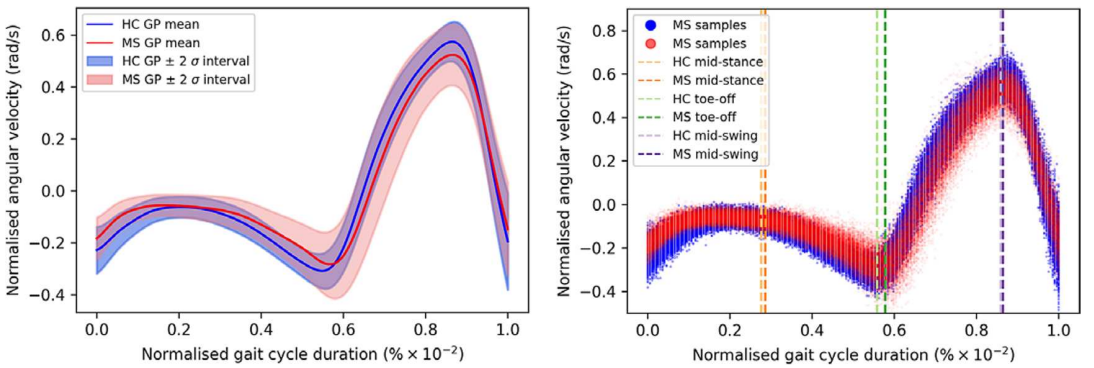

Figure 6. *Left: GP group predictions, Right: Samples drawn from the GPs. The vertical dotted lines, from left to right, correspond to the group-averaged mid-stance, toe-off and mid-swing events. Toe-off events were defined as the time point where the minimum negative peak occurs immediately before the maximum positive peak during each stride. The mid-stance point corresponds to the halfway point between the start of the gait cycle and the toe-off event, whereas the mid-swing corresponds to the maximum amplitude point in the signal.*

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

using samples drawn from the predictive distributions of both groups, as described in Equation 13. In this context, $\mathbb{P}$ and $\mathbb{Q}$ represent the posterior distributions under examination, corresponding to the HC and MS groups, respectively. p_i and q_i denote samples drawn from these distributions, while m and n indicate the respective sample sizes of $\mathbb{P}$ and $\mathbb{Q}$. The benefit of using the MMD is that it is a nonparametric distance metric employing kernel embeddings of the distributions considered for comparison. This presents an advantage, as the kernel trick enables the efficient assessment of effectively infinite moments by employing inner products within a feature space (Bishop, 2006). While the Kullback–Leibler (KL) divergence between univariate Gaussian distributions is analytically available in closed form, the MMD was selected as the primary metric for this analysis based on two key methodological considerations. First, while designating the HC group as the reference distribution in KL calculations would be a natural choice for clinical interpretation, MMD's symmetric property eliminates this decision requirement altogether. Second, and more significantly, as demonstrated in Appendix E, MMD and KL divergence quantify fundamentally different aspects of distributional differences. MMD correlates more directly with the difference process between the MS and HC distributions, with sensitivity to the combined variance of both distributions. In contrast, KL divergence depends primarily on the ratio of variances, making it disproportionately sensitive to small values in the reference distribution rather than capturing the absolute magnitude of differences. This mathematical property explains why MMD provides a more balanced assessment of group-level gait pattern differences, especially in regions where the variance of the difference process is high. Specific to the MMD computation, the radial basis function (RBF) (Lloyd and Ghahramani, 2015) has been used as the preferred kernel, defined according to Equation 14, where the σ parameter controlling the bandwidth of the kernel has been set as the median distance between points in the aggregate sample (Gretton et al., 2012).

$$MMD^2(\mathbb{P},\mathbb{Q}) = \frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j\neq i}^{m} k(p_i,p_j) + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} k(q_i,q_j) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(p_i,q_j) \tag{13}$$

$$k(p,q) = \exp\left(-\frac{\|p-q\|^2}{2\sigma^2}\right) \tag{14}$$

The predictive distributions are visible on the left of Figure 6. On the right side, 500 samples have been generated at each of the 200 equidistant test points along the input space. The comparison is depicted in Figure 7, where the values of the MMD test statistic have been normalized between 0 and 1. While the interpretation of the absolute values of the MMD is difficult, here, values closer to 0 indicate similarity between distributions, whereas increasingly higher values are indicative of greater discrepancies between the models. Moreover, to aid comparison, absolute differences in terms of mean predictions and standard deviations are also shown, where it should be noted that the $\pm 2\sigma$ interval corresponds to the standard deviation of the difference process and should not be confused with the bandwidth hyperparameter of the MMD. Even though certain regions in the gait cycle exhibit notable similarities (indicated by the black and dark blue regions), specific locations stand out as being dissimilar. These regions include the first 15%, the range of 35 to 55%, and the range of 60 to 90% of the gait cycle. The first two regions in which discrepancies have been highlighted overlap with the double support phases (that is the period in which both feet are in contact with the ground), (Neumann, 2016), whereas the greatest discrepancies are most apparent during the initiation of the swing phase.

Direct comparison with other results presented in the relevant literature is challenging, as most of the characterizations of pathological gait primarily rely on joint kinematics, instead of segment kinematics. Nevertheless, in the case of the MS group, it is evident that the variability in the gait pattern progressively increases from mid-stance to mid-swing. Consistent with the findings of this study, other researchers have also reported greater variability in joint angles among individuals with MS, even among those with mild disability (Crenshaw et al., 2006; Kelleher et al., 2010; Nogueira et al., 2013; Severini et al. 2017). Similar

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

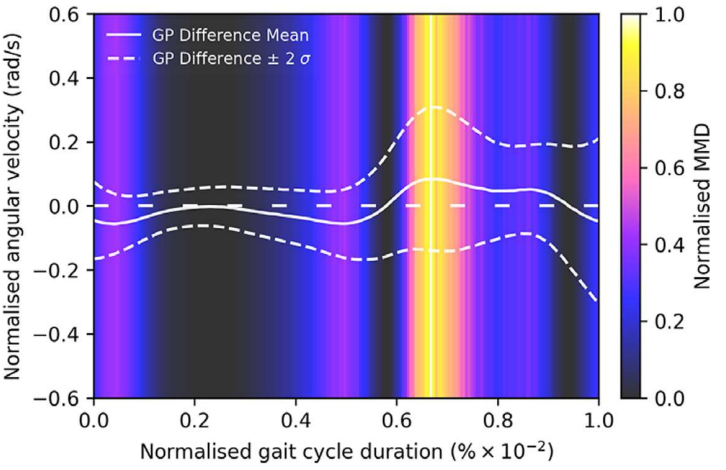

Figure 7. *Visualization of the group differences. The solid white line shows the difference between the means of the two GPs, whereas the curved dashed lines showcase two standard deviations from the mean. The horizontal dashed line has been added here only for highlighting zero-crossing points. Notably, although the* $\pm 2\sigma$ *interval may not seem symmetric above and below the mean upon initial observation, it is, in fact, symmetric. Disparities between the HC and MS models are highlighted across the input space using the MMD.*

trends have been found by Salehi et al. (2020), where the authors reported significant differences between HC and MS during the stance and swing phase when looking at the deviation phase, as a measure of coordination variability. Additionally, the flatter trajectory observed during the stance phase may suggest a reduced range of plantar flexion, as reported by Nogueira et al. (2013). This reduction can lead to diminished power generation at the ankle, particularly during the initiation of the swing phase, potentially impacting the stability control throughout this latter phase. Such outcomes might be attributed to muscle weakness, spasticity, fatigue, or balance impairments (Nogueira et al., 2013; Socie et al., 2013). Therefore, while confirming authors' expectations, these group-level comparative results highlight the model's ability to effectively capture the inherent variability of MS gait patterns, allowing for an informed analysis.

Next, to showcase the capabilities of the proposed hierarchical modeling approach to facilitate nuanced comparisons at lower levels in the hierarchy, individual gait pattern comparisons have been investigated, relative to the control group. This analysis aligns with the overarching aim of providing a patient-specific assessment. As such, detailed shank angular velocity predictions were generated for all participants in the study, which can be seen in Figure 8. Here, the held-out test data has been overlaid, allowing for enhanced insight into the model's predictive performance. For clarity, the overlaid data corresponds to 30% of randomly selected datapoints from each particular individual, which were not seen during training. The individual NMSE and MSLL scores are provided in Appendix F. Notably, individuals with MS exhibited significantly higher NMSE scores, indicating lower predictive capabilities of the MS model. Not surprisingly, the MS group also displayed significantly higher MSLL scores, further supporting the diminished predictive performance of the model for this population, given the increased variability in the gait patterns as well as the presence of asymmetry. To highlight individual-level discrepancies, samples drawn from the HC group predictive distribution were compared to samples drawn from the individual GPs, in a similar way to the MMD-based approach used above. More specifically, 500 samples were generated again at each of the 200 equidistant test locations along the input space, and the MMD was computed at every test location, for every subject. Then, the MMD values from all subjects were aggregated and normalized between 0 and 1. Therefore, the regions in the gait cycle which are most different relative to the control group are highlighted in Figure 9. As a result, while many of the PwMS

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

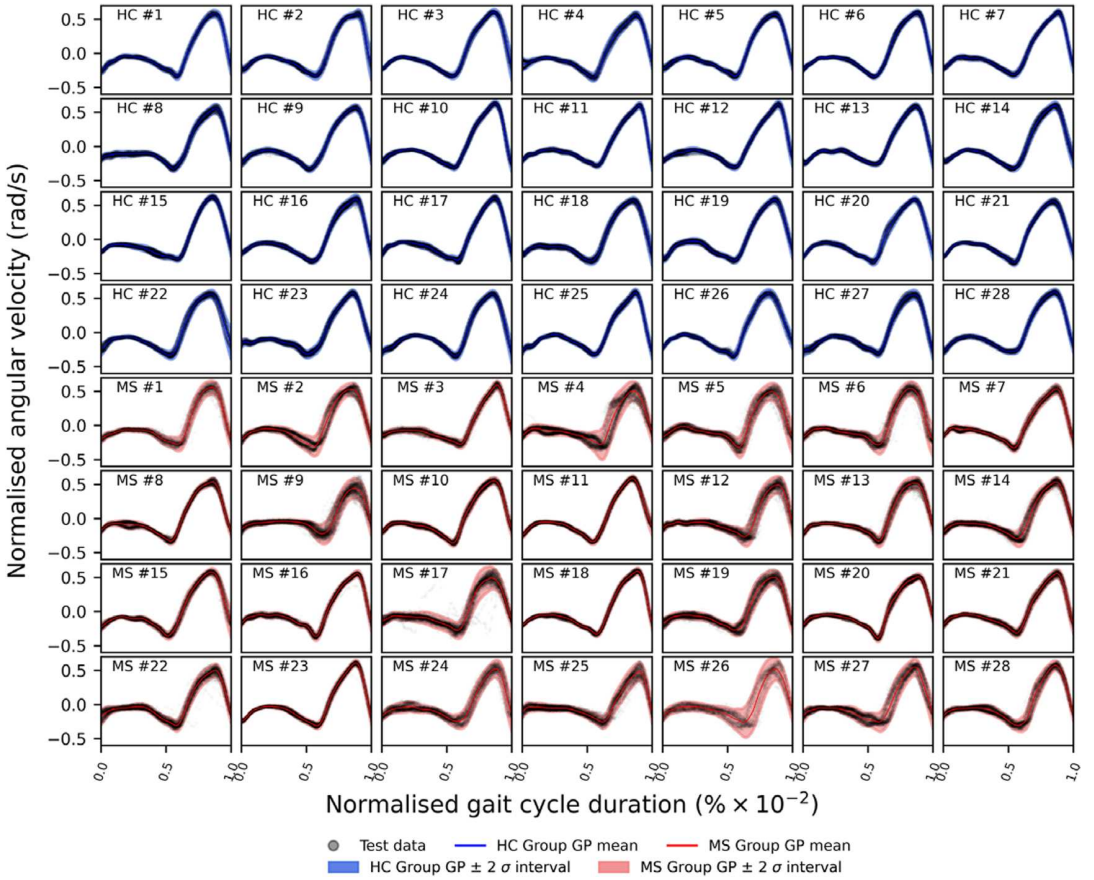

Figure 8. *Individual predictions versus held-out test data. The first four rows correspond to HC individuals, while the last four rows correspond to PwMS.*

appear to display a walking pattern closely aligned with the HC group pattern, with no immediate visible signs of impairment, several individuals with MS clearly stand out. However, it is important to acknowledge that comparing the individual MS predictions with the predictive distribution of the HC group is not sufficient. This is due to the presence of overlapping confidence intervals and comparable means shared between the HC group and the PwMS, stemming from the increased variability in the individual predictions of the latter group.

Figure 10 highlights the individual heteroscedastic variances, where departures from the contained HC gait patterns are evident for several individuals affected by MS, especially toward the termination of the stance and during the swing phase. Therefore, extending the hierarchical methodology to model individual gait patterns led to a novel feature that may serve as an indicator of the presence of neurological or musculoskeletal deficits. Two main factors are believed to be responsible for these results. The first one might be associated with the possible lack of control during the swing phase, as a result of muscle spasticity, fatigue or balance impairments (Nogueira et al., 2013; Socie et al., 2013), while the second one pertains to the presence of asymmetry. For clarification, the alternative concept of asymmetry used here will be elaborated upon in the subsequent paragraphs, where the hierarchical model has been extended to include separate consideration for the left and right limbs. This extension has been added since the distribution over the outputs must be symmetric about the mean, which is not representative of the bi-modal distribution when asymmetry is present. However, the asymmetric gait patterns are still captured within the $\pm 2\sigma$ bounds of the GP predictions. Thus, the variance predicted throughout the gait cycle could

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

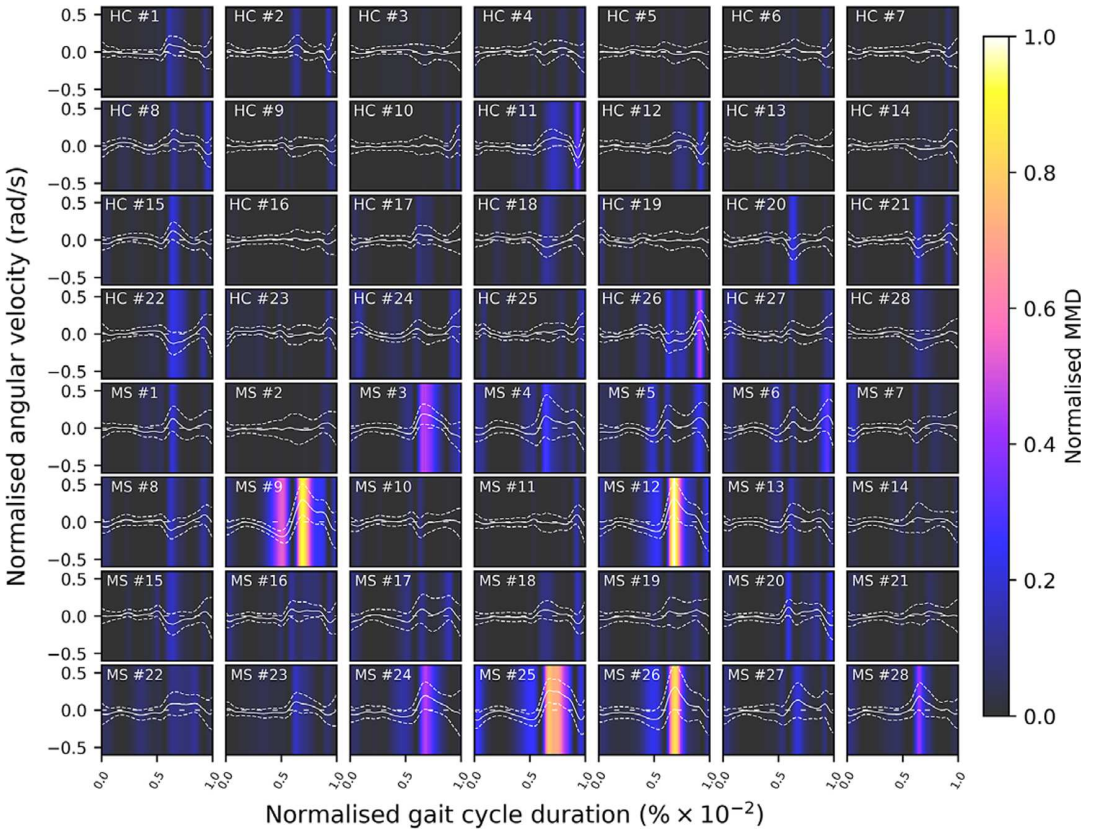

Figure 9. *Individual differences, MMD highlighted. The first four rows correspond to HC individuals, while the last four rows correspond to PwMS.*

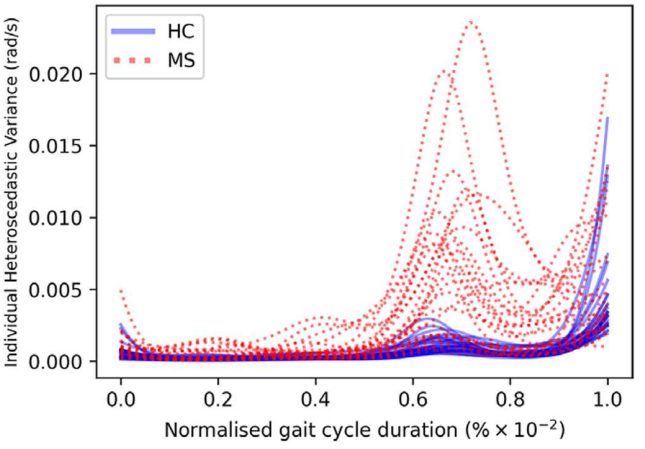

Figure 10. *Heteroscedastic variance differences at the individual layer in the hierarchy, where data from the left and right limbs are combined for each individual. Each line corresponds to a single subject.*

hold significant potential for integration into clinical gait analysis, noting that our methodology is not intended as a substitute for standard gait analysis, but rather as a valuable augmentation to the clinical assessment of patients with MS.

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

3.4. Quantifying individual asymmetry

In light of the results presented in Section 3.3, where the heteroscedastic variance has been observed to be one of the discriminant factors between HCs and PwMS, as well as considering the data normalization procedure, this section addresses the third objective of the case study concerning a novel proposal for asymmetry quantification. Recognizing asymmetry as one of the key factors for the increased variability in the gait patterns, the hierarchical model has been extended to separately address contralateral limbs. This extension not only allows for a nuanced exploration of contralateral limb data but also serves as a foundational step in introducing a novel methodology for quantifying lower limb asymmetry from a probabilistic perspective.

Traditionally, gait asymmetry has been defined as the absolute difference between the temporal metrics extracted from the left and right lower limbs (Godfrey et al., 2015) or as the natural logarithm of the absolute ratio of the shorter and the longer mean value of the temporal metric (Yogev et al., 2007). Gait asymmetry is often considered to be an indicator of MS (Angelini et al., 2021). However, in the context of this work, the term asymmetry is used to denote signals that align for the most part but display disparities at certain locations. To allow improved visual insights into the presence of asymmetry, the individual left and right model predictions are presented in Figure 11, relative to the unseen held-out test data.

Analyzing Figure 11, disparities between left and right limbs for the MS-affected subject are evident, and several subjects clearly stand out. Conversely, the gait patterns observed for the HCs demonstrate a higher degree of symmetry, that is the same functional form throughout the gait cycle, despite the sparse presence of slight deviations between the left and right shanks (which can be considered negligible). Given the normalization procedure employed in this study, traditional asymmetry metrics can no longer be computed, necessitating the adoption of an alternative, comprehensive methodology. As such, first, the Wasserstein distance (Villani, 2009) between samples drawn from each of the left and right limb GPs was computed across the input space. By computing the Wasserstein distance (WD) between these samples, a detailed insight into gait asymmetry can be obtained, effectively measuring how much the left and right limbs differ across the entire gait cycle. Similarly to the MMD distance, 500 samples were drawn from each GP's predictive distributions at 100 equidistant points spread across the input space. Then, the WD was computed at each of the test locations (please refer to Figure G1 in Appendix G). This allows the end-users to visually discern the distinctive locations within the gait cycle where asymmetry is more prominent. Moreover, the overall effect of gait asymmetry was quantified by computing the area under the curve, thus providing a unified gait asymmetry metric. The comparison of the individual differences for all subjects included in the study can be seen in Figure 12a,b.

Here, the PwMS numbered 2, 4, 19, 26, and 27 particularly stand out. Subject 2 displayed a reduced range of mobility on the right shank during the stance phase, as well as an increased stance duration, when compared to the left leg. An uncontrolled movement was recorded on the right leg of subject 4. For this subject, the one-sided balance and coordination deficits were evident across the entire gait cycle, albeit more pronounced during the swing phase. Subject 19 also displayed temporal differences between the left and right limbs. A reduced range of mobility was also recorded for subject 26 for the right shank. This was recorded in conjunction with temporal differences between the two limbs, as well as increased variability during the swing phase and the end of the double support phase. Finally, subject 27 exhibited an uncontrolled right shank movement, accompanied by temporal differences as well as higher movement variability on the left leg. To summarize the asymmetry differences between the HC and MS groups, the nonparametric Mann–Whitney U test was employed, with a minimum significance alpha level of 5%, highlighting statistically significant differences between the two groups ($p < 0.0001$, see Figure 12c). Nonetheless, the correlation between the WD and the clinical scores (in the form of the EDSS score) was also explored. Only a weak correlation has been found ($Pearson's\, r = 0.33$), which might suggest that asymmetry could manifest independently of disease severity in MS. Despite some arguments proposing a link between asymmetry and disease severity, such as Pau et al. (2021), which demonstrated moderate correlations between joint kinematics asymmetry and the EDSS score using trend symmetry (Crenshaw and Richards, 2006), the findings of this study emphasize asymmetry relevance across various levels of

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

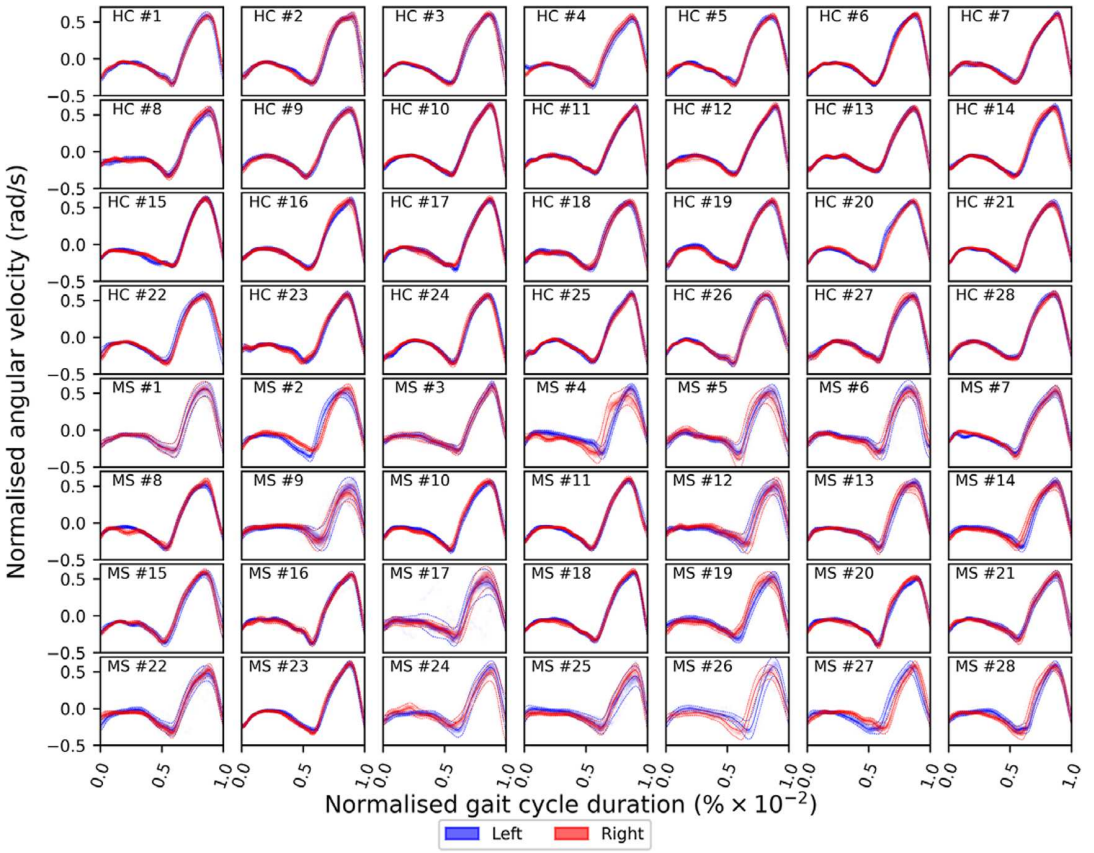

Figure 11. *Individual limb GP predictions. The first four rows correspond to HCs, while the last four rows correspond to PwMS. The solid lines represent the mean GP predictions, while the dotted lines correspond to the $\pm 2\sigma$ interval. The dots represent the held-out test data.*

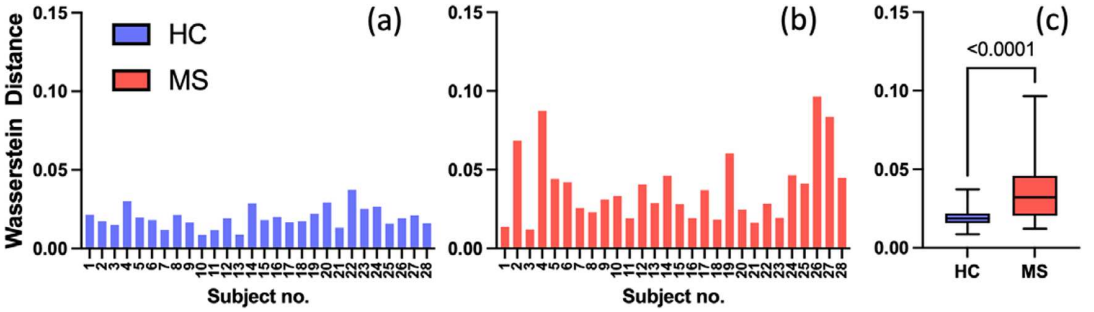

Figure 12. *(a) HC, (b) MS—Unified Wasserstein distance computed between samples drawn from the left GP and the right GP for each of the individuals in both HC and MS groups. (c) Statistical comparison between the HC and MS groups.*

MS disability. However, because of the considerably lower number of subjects included in the present study, drawing definitive clinical conclusions is not advisable.

Second, an additional dimension of asymmetry analysis was pursued using the KL divergence. This process involved comparing the third-layer GP model that considers both limbs together with separate

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

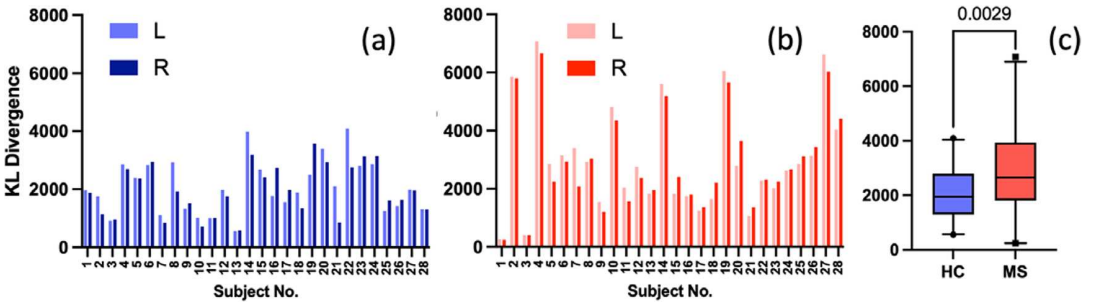

Figure 13. *KL divergence computed between the individual level GP, combining the left and right limbs (third layer), and the individual limbs GP models, treating each limb individually (fourth layer). (a) HC, (b) MS. The lighter shade represents the left shank, while the darker shade represents the right shank. (c) statistical comparison.*

fourth-layer GP models that focus on the left and right limbs individually. This can be visualized in Figure 13a,b. The KL divergence, therefore, gives an indication of how well the combined model represents individual limb dynamics. A high KL divergence indicates that the combined model might not fully capture the nuances of each of the individual limbs, potentially reflecting asymmetry or distinctive characteristics. It should be noted that computing the KL divergence involves treating the GPs as multivariate Gaussian distributions and requires access to the full covariance matrix. In contrast to the previous asymmetry analysis employing the WD, analyzing the KL divergence metrics in Figure 13b, two additional MS subjects stand out, namely subjects 10 and 14. Although outcomes of subject 10 are not readily apparent, this result can be attributed to a subtle discrepancy between the left and right gait trajectories, which may elude visual inspection. Similar results were also recorded for subject 14, for which the presence of asymmetry was significantly more pronounced. The statistical comparison is presented in Figure 13c. The nonparametric Mann–Whitney U test was also employed in this case, with a minimum significance alpha level of 5%, highlighting statistically significant differences between the two groups ($p = 0.0029$). However, no correlations between the KL divergence-based asymmetry metrics and EDSS score were found (*Pearson's* $r = 0.1015$ and 0.0051 for the left and right shanks respectively), necessitating additional validation procedures before drawing any definitive clinical implications. However, the extension of the hierarchical model to individually consider contralateral limbs provided new insights into asymmetry levels in individuals affected by MS through a novel, multifaceted approach employing both the WD and the KL divergence as complementary asymmetry metrics. It is important to note that while the results presented here may suggest that asymmetry could be a challenging aspect of MS even in the early disease stages, caution should be exercised due to the small sample size and the need for further validation. Nonetheless, these insights may hold promise in assisting end-users toward providing improved personalized treatment or rehabilitation plans.

While the advantages and potential applications of the newly proposed methodology for characterizing the shank angular velocity for PwMS have been emphasized, it is also important to consider its potential limitations. First, while the size of the inducing point set has been kept fixed and shared across all levels of the proposed hierarchy, in order to reduce the parameter space over which to optimize, future work should investigate how this design constraint affects the quality of the sparse approximation within the context of this work. Second, while the authors retained the entire dataset for the analysis proposed in this study, future work may also investigate the effects of downsampling and attempting the modeling problem with a reduced sample size. This may allow direct comparison of sparse GP approximations to standard GP regression.

In summary, the contribution of this work has been to propose a novel methodology for investigating similarities and differences in gait data across multiple hierarchical levels, presenting an initial case study of HVSHGPs directly applied to gait signals. The proposed approach is to make a departure from

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

understanding gait with respect to a set of summary features; instead, the full gait cycle is modeled functionally, including varying noise/uncertainty throughout the signal. This model additionally obeys the underlying hierarchical structure of the gait data, exploiting similarities, but respecting differences. Finally, given that the shank angular velocity can be directly measured using wearable sensors—a cost-effective and versatile technology applicable in both laboratory and free-living environments—this signal presents a promising candidate for a biomarker in longitudinal studies of distal lower limb motion, offering advantages over traditional motion capture systems.

4. Conclusions

To the best of our knowledge, this study marks the first instance where scalable hierarchical GPs have been employed as a flexible Bayesian machine learning technique for modeling the nonlinear shank angular velocity. First, the hierarchical approach has been used to account for the temporally structured covariance between groups of patients and individual subjects. Then, the problem of scaling GPs to handle large datasets, such as the ones obtained during clinical gait assessments, has been addressed by using variational approximation methods. Finally, the variability in the gait cycle has been captured by input-dependent noise modeling, that is heteroscedasticity. This was achieved by modeling the log-noise variance of the process as an additional GP, which is also learnt in a variational manner. In possession of this hierarchical model is then possible to perform comparisons of the gait cycle across the full function space, rather than simply on a discrete feature level.

In conclusion, the move toward probabilistic modeling of the shank angular velocity allows for the use of probabilistic models, which offer robust and accurate mean predictions, along with automatic uncertainty estimates. The use of these models facilitates a better understanding of the impaired gait pattern in MS and also holds promise for possible extension to other pathological gait conditions.

Data availability statement. De-identified data underlying this study may be made available upon reasonable request to the corresponding author, subject to review approval by the relevant data access committees and under a suitable data sharing agreement.

Acknowledgments. 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.

Author contribution. Conceptualization: A.S., T.J.R., E.J.C., C.M.; Methodology: A.S, T.J.R.; Data curation: A.S.; Data visualization: A.S.; Writing original draft: A.S., T.J.R., E.J.C., C.M.; All authors approved the final submitted draft.

Funding statement. This study was funded by the National Institute for Health Research (NIHR) through the Sheffield Biomedical Research Centre (BRC, grant number IS-BRC-1215-20017) and PhD scholarship Grant no. 820820. Authors E.J. Cross and T.J. Rogers were supported by EP/W005816/1. T.J. Rogers is additionally supported by EP/W002140/1.

Competing interests. The authors declare none.

Ethical standard. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

Abadi M, **Agarwal A**, **Barham P**, **Brevdo E**, **Chen Z**, **Citro C**, **Corrado GS**, **Davis A**, **Dean J**, **Devin M**, **Ghemawat S**, **Goodfellow I**, **Harp A**, **Irving G**, **Isard M**, **Jia Y**, **Jozefowicz R**, **Kaiser L**, **Kudlur M**, **Levenberg J**, **Mane D**, **Monga R**, **Moore S**, **Murray D**, **Olah C**, **Schuster M**, **Shlens J**, **Steiner B**, **Sutskever I**, **Talwar K**, **Tucker P**, **Vanhoucke V**, **Vasudevan V**, **Viegas F**, **Vinyals O**, **Warden P**, **Wattenberg M**, **Wicke M**, **Yu Y and Zheng X** (2015) TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org.

Anderson FC and Pandy MG (2001) Dynamic optimization of human walking. *Journal of Biomechanical Engineering 123*(5), 381–390.

Angelini L, **Hodgkinson W**, **Smith C**, **Dodd JM**, **Sharrack B**, **Mazzà C and Paling D** (2020) Wearable sensors can reliably quantify gait alterations associated with disability in people with progressive multiple sclerosis in a clinical setting. *Journal of Neurology 267*(10), 2897–2909.

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

Angelini L, **Buckley E**, **Bonci T**, **Radford A**, **Sharrack B**, **Paling D**, **Nair KPS and Mazza C** (2021) A multifactorial model of multiple sclerosis gait and its changes across different disability levels. *IEEE Transactions on Biomedical Engineering 68*(11), 3196–3204.

Arcolin I, **Corna S**, **Giardini M**, **Giordano A**, **Nardone A and Godi M** (2019) Proposal of a new conceptual gait model for patients with Parkinson's disease based on factor analysis. *Biomedical Engineering Online 18*(1), 1–18.

Benemerito I, **Montefiori E**, **Marzo A and Mazzà C** (2022) Reducing the complexity of musculoskeletal models using Gaussian process emulators. *Applied Sciences (Switzerland) 12*, 12932.

Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York: Springer

Bui TD, **Yan J and Turner RE** (2017) A unifying framework for Gaussian process pseudo-point approximations using power expectation propagation. *Journal of Machine Learning Research 18*, 1–72.

Chen Z, **Guo Q**, **Li T**, **Yan Y and Jiang D** (2023) Gait prediction and variable admittance control for lower limb exoskeleton with measurement delay and extended-state-observer. *IEEE Transactions on Neural Networks and Learning Systems 34*(11), 8693–8706.

Chun C, **Kim S-J**, **Hong J and Park FC** (2015) Gaussian process learning and interpolation of gait motion for rehabilitation robots. *2015 6th International Conference on Automation, Robotics and Applications (ICARA)* pp. 198–203.

Cicirelli G, **Impedovo D**, **Dentamaro V**, **Marani R**, **Pirlo G and D'Orazio TR** (2022) Human gait analysis in neurodegenerative diseases: A review. *IEEE Journal of Biomedical and Health Informatics 26*, 229–242.

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

Cortes C and Vapnik V (1995) Support-vector networks. *Machine Learning 20*(3), 273–297.

Creagh AP, **Dondelinger F**, **Lipsmeier F**, **Lindemann M and De Vos M** (2022) Longitudinal trend monitoring of multiple sclerosis ambulation using smartphones. *IEEE Open Journal of Engineering in Medicine and Biology 3*, 202–210.

Crenshaw SJ and Richards JG (2006) A method for analyzing joint symmetry and normalcy, with an application to analyzing gait. *Gait and Posture 24*(4), 515–521.

Crenshaw SJ, **Royer TD**, **Richards JG and Hudson DJ** (2006) Gait variability in people with multiple sclerosis. *Multiple Sclerosis Journal 12*(5), 613–619.

de G. Matthews AG, **van der Wilk M**, **Nickson T**, **Fujii K**, **Boukouvalas A**, **León-Villagrá P**, **Ghahramani Z and Hensman J** (2017) Gpflow: A Gaussian process library using tensorflow. *Journal of Machine Learning Research 18*(40), 1–6.

Dozat T (2016) Incorporating Nesterov momentum into Adam. *Proceedings of the 4th International Conference on Learning Representations*, pp. 1–4.

Fang B, **Zhou Q**, **Sun F**, **Shan J**, **Wang M**, **Xiang C and Zhang Q** (2020) Gait neural network for human-exoskeleton interaction. *Frontiers in Neurorobotics 14*(October), 1–9.

Filli L, **Sutter T**, **Easthope CS**, **Killeen T**, **Meyer C**, **Reuter K**, **Lörincz L**, **Bolliger M**, **Weller M**, **Curt A**, **Straumann D**, **Linnebank M and Zörner B** (2018) Profiling walking dysfunction in multiple sclerosis: Characterisation, classification and progression over time. *Scientific Reports 8*(1), 1–13.

Gadaleta, M., **Merelli, L., and Rossi, M.** (2016). Human authentication from ankle motion data using convolutional neural networks. *IEEE Workshop on Statistical Signal Processing Proceedings* 2016 (August), 1–5.

Gadaleta M and Rossi M (2018) IDNet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognition 74*, 25–37.

Gelfand, J. M. (2014). Chapter 12 – Multiple sclerosis: Diagnosis, differential diagnosis, and clinical presentation. In Goodin DS (ed), *Multiple Sclerosis and Related Disorders*, volume 122 of *Handbook of Clinical Neurology.* Elsevier, pp. 269–290.

Gil-Castillo J, **Alnajjar F**, **Koutsou A**, **Torricelli D and Moreno JC** (2020) Advances in neuroprosthetic management of foot drop: A review. *Journal of Neuroengineering and Rehabilitation 17*, 1–19.

Glackin C, **Salge C**, **Greaves M**, **Polani D**, **Slavnić S**, **Ristić-Durrant D**, **Leu A and Matjačić Z** (2014). Gait trajectory prediction using Gaussian process ensembles. *IEEE-RAS International Conference on Humanoid Robots*, pp. 628–633.

Godfrey A, **Del Din S**, **Barry G**, **Mathers JC and Rochester L** (2015) Instrumenting gait with an accelerometer: A system and algorithm examination. *Medical Engineering and Physics 37*(4), 400–407.

Gretton A, **Borgwardt KM**, **Rasch MJ**, **Schölkopf B and Smola A** (2012) A kernel two-sample test. *Journal of Machine Learning Research 13*, 723–773.

Haji Ghassemi N, **Hannink J**, **Martindale CF**, **Gaßner H**, **Müller M**, **Klucken J and Eskofier BM** (2018) Segmentation of gait sequences in sensor-based movement analysis: A comparison of methods in Parkinson's disease. *Sensors 18*, 145.

Hausdorff JM, **Peng C-K**, **Ladin ZVI**, **Wei JY and Goldberger AL** (1995) Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. *Journal of Applied Physiology 78*(1), 349–358.

Hensman J, **Durrande N and Solin A** (2018) Variational fourier features for gaussian processes. *Journal of Machine Learning Research 18*(151), 1–52.

Hensman J, **Lawrence ND and Rattray M** (2013) Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. *BMC Bioinformatics 14*(1), 1–12.

Hong J, **Chun C**, **Kim S-J and Park FC** (2019) Gaussian process trajectory learning and synthesis of individualized gait motions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 27*(6), 1236–1245.

Horst F, **Kramer F**, **Schäfer B**, **Eekhoff A**, **Hegen P**, **Nigg BM and Schöllhorn WI** (2016) Daily changes of individual gait patterns identified by means of support vector machines. *Gait and Posture 49*, 309–314.

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

Horst F, **Lapuschkin S**, **Samek W**, **Müller K-R and Schöllhorn WI** (2019) Explaining the unique nature of individual gait patterns with deep learning. *Scientific Reports 9*(1), 1–13.

Ingelse L, **Branco D**, **Gjoreski H**, **Guerreiro T**, **Bouca-Machado R**, **Ferreira JJ and CNS Physiotherapy Study Group** (2022) Personalised gait recognition for people with neurological conditions. *Sensors 22*(11), 3980.

Kelleher KJ, **Spence WD**, **Solomonidis S and Apatsidis D** (2010) The effect of textured insoles on gait patterns of people with multiple sclerosis. *Gait and Posture 32*(1), 67–71.

Kurtzke JF (1983) Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology 33*, 1444–1452.

Lázaro-Gredilla, M. and Titsias, M. K. (2011). Variational heteroscedastic Gaussian process regression. In *Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11*. Madison, WI: Omnipress, pp. 841–848.

Liu H, **Ong YS and Cai J** (2018) Large-scale heteroscedastic regression via Gaussian process. *IEEE Transactions on Neural Networks and Learning Systems 32*(2), 708–721.

Lloyd JR and Ghahramani Z (2015) Statistical model criticism using kernel two sample tests. *Advances in Neural Information Processing Systems 28*.

Lord S, **Galna B**, **Verghese J**, **Coleman S**, **Burn D and Rochester L** (2013) Independent domains of gait in older adults and associated motor and nonmotor attributes: Validation of a factor analysis approach. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences 68*(7), 820–827.

Michalis, T. (2009). Variational model selection for sparse Gaussian process regression. Technical report, School of Computer Science, University of Manchester.

Moe-Nilssen R (1998) A new method for evaluating motor control in gait under real-life environmental conditions. Part 1: The instrument. *Clinical Biomechanics 13*, 320–327.

Moon Y, **Sung JH**, **An R**, **Hernandez ME and Sosnoff JJ** (2016) Gait variability in people with neurological disorders: A systematic review and meta-analysis. *Human Movement Science 47*, 197–208.

Moon Y, **McGinnis RS**, **Seagers K**, **Motl RW**, **Sheth N**, **Wright Jr JA**, **Ghaffari R and Sosnoff JJ** (2017) Monitoring gait in multiple sclerosis with novel wearable motion sensors. *PLoS One 12*(2), 1–19.

Neptune RR, **Clark DJ and Kautz SA** (2009) Modular control of human walking: A simulation study. *Journal of Biomechanics 42*(9), 1282–1287.

Neumann DA (2016) *Kinesiology of the Musculoskeletal System: Foundations for Rehabilitation*. St. Louis, Missouri, US: Elsevier Health Sciences, pp. 627–681.

Nogueira LAC, **Teixeira L**, **Sabino P**, **Filho HA**, **Alvarenga RMP and Thuler LC** (2013) Gait characteristics of multiple sclerosis patients in the absence of clinical disability. *Disability and Rehabilitation 35*(17), 1472–1478.

Panebianco GP, **Bisi MC**, **Stagni R and Fantozzi S** (2018) Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements. *Gait and Posture 66*, 76–82.

Pasciuto I, **Bergamini E**, **Iosa M**, **Vannozzi G and Cappozzo A** (2017) Overcoming the limitations of the harmonic ratio for the reliable assessment of gait symmetry. *Journal of Biomechanics 53*, 84–89.

Pau M, **Mandaresu S**, **Pilloni G**, **Porta M**, **Coghe G**, **Marrosu MG and Cocco E** (2017) Smoothness of gait detects early alterations of walking in persons with multiple sclerosis without disability. *Gait and Posture 58*, 307–309.

Pau M, **Leban B**, **Deidda M**, **Putzolu F**, **Porta M**, **Coghe G and Cocco E** (2021) Kinematic analysis of lower limb joint asymmetry during gait in people with multiple sclerosis. *Symmetry 13*, 598.

Polhemus A, **Delgado-Ortiz L**, **Brittain G**, **Chynkiamis N**, **Salis F**, **Gaßner H**, **Gross M**, **Kirk C**, **Rossanigo R**, **Taraldsen K**, **Balta D**, **Breuls S**, **Buttery S**, **Cardenas G**, **Endress C**, **Gugenhan J**, **Keogh A**, **Kluge F**, **Koch S**, **Micó-Amigo ME**, **Nerz C**, **Sieber C**, **Williams P**, **Bergquist R**, **de Basea MB**, **Buckley E**, **Hansen C**, **Mikolaizak AS**, **Schwickert L**, **Scott K**, **Stallforth S**, **van Uem J**, **Vereijken B**, **Cereatti A**, **Demeyer H**, **Hopkinson N**, **Maetzler W**, **Troosters T**, **Vogiatzis I**, **Yarnall A**, **Becker C**, **Garcia-Aymerich J**, **Leocani L**, **Mazzà C**, **Rochester L**, **Sharrack B**, **Frei A**, **Puhan M and Mobilise D** (2021) Walking on common ground : A cross-disciplinary scoping review on the clinical utility of digital mobility outcomes. *NPJ Digital Medicine 4*(1), 149.

Quiñonero-Candela J and Rasmussen CE (2005) A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research 6*(65), 1939–1959.

Rasmussen CE and Ghahramani Z (2001) Occam's razor. *Advances in Neural Information Processing Systems*, 294–300.

Rasmussen CE and Williams CK (2006) *Gaussian Processes for Machine Learning*. The MIT Press

Rasouli, F. and Reed, K. B. (2021). Identical limb dynamics for unilateral impairments through biomechanical equivalence. *Symmetry 13*(4), 705.

Remelius JG, **Jones SL**, **House JD**, **Busa MA**, **Averill JL**, **Sugumaran K**, **Kent-Braun JA and Van Emmerik RE** (2012) Gait impairments in persons with multiple sclerosis across preferred and fixed walking speeds. *Archives of Physical Medicine and Rehabilitation 93*(9), 1637–1642.

Rodríguez-Martín D, **Samà A**, **Pérez-López C**, **Català A**, **Moreno Arostegui JM**, **Cabestany J**, **Bayés A**, **Alcaine S**, **Mestre B**, **Prats A**, **Cruz Crespo M**, **Counihan TJ**, **Browne P**, **Quinlan LR**, **ÓLaighin G**, **Sweeney D**, **Lewy H**, **Azuri J**, **Vainstein G**, **Annicchiarico R**, **Costa A and Rodríguez-Molinero A** (2017) Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer. *PLoS One 12*(2), 1–26.

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

Rogers TJ, **Gardner P**, **Dervilis N**, **Worden K**, **Maguire AE**, **Papatheou E and Cross EJ** (2020) Probabilistic modelling of wind turbine power curves with application of heteroscedastic Gaussian process regression. *Renewable Energy 148*, 1124–1136.
Rueterbories J, **Spaich EG**, **Larsen B and Andersen OK** (2010) Methods for gait event detection and analysis in ambulatory systems. *Medical Engineering & Physics 32*(6), 545–552.
Salehi R, **Mofateh R**, **Mehravar M**, **Negahban H**, **Tajali S and Monjezi S** (2020) Comparison of the lower limb inter-segmental coordination during walking between healthy controls and people with multiple sclerosis with and without fall history. *Multiple Sclerosis and Related Disorders 41*, 102053.
Severini G, **Manca M.**, **Ferraresi G**, **Caniatti LM**, **Cosma M**, **Baldasso F**, **Straudi S**, **Morelli M and Basaglia N** (2017) Evaluation of clinical gait analysis parameters in patients affected by multiple sclerosis: Analysis of kinematics. *Clinical biomechanics 45*, 1–8.
Shema-Shiratzky S, **Gazit E**, **Sun R**, **Regev K**, **Karni A**, **Sosnoff JJ**, **Herman T**, **Mirelman A and Hausdorff JM** (2019) Deterioration of specific aspects of gait during the instrumented 6-min walk test among people with multiple sclerosis. *Journal of Neurology 266*, 3022–3030.
Smola AJ and Schölkopf B (2004) A tutorial on support vector regression. *Statistics and Computing 14*, 199–222.
Socie MJ, **Sandroff BM**, **Pula JH**, **Hsiao-Wecksler ET**, **Motl RW and Sosnoff JJ** (2013) Footfall placement variability and falls in multiple sclerosis. *Annals of Biomedical Engineering 41*(8), 1740–1747.
Titsias MK (2009) Variational learning of inducing variables in sparse Gaussian processes. *Journal of Machine Learning Research 5*, 567–574.
Verghese J, **Wang C**, **Lipton RB**, **Holtzer R and Xue X** (2007) Quantitative gait dysfunction and risk of cognitive decline and dementia. *Journal of Neurology, Neurosurgery and Psychiatry 78*(9), 929–935.
Verghese J, **Holtzer R**, **Lipton RB and Wang C** (2009) Quantitative gait markers and incident fall risk in older adults. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences 64*(8), 896–901.
Villani C (2009) *The Wasserstein Distances*. Berlin, Heidelberg: Springer, pp. 93–111
Wang JM, **Fleet DJ and Hertzmann A** (2008) Gaussian process dynamical models for human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence 30*(2), 283–298.
Winner TS, **Rosenberg MC**, **Jain K**, **Kesar TM**, **Ting LH and Berman GJ** (2023) Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics. *PLoS Computational Biology 19*(10), 1–33.
Wu X, **Liu D-X**, **Liu M**, **Chen C and Guo H** (2018) Individualized gait pattern generation for sharing lower limb exoskeleton robot. *IEEE Transactions on Automation Science and Engineering 15*(4), 1459–1470.
Xiang Y, **Arora JS and Abdel-Malek K** (2010) Physics-based modeling and simulation of human walking: A review of optimization-based and other approaches. *Structural and Multidisciplinary Optimization 42*, 1–23.
Yogev G, **Plotnik M**, **Peretz C**, **Giladi N and Hausdorff JM** (2007) Gait asymmetry in patients with Parkinson's disease and elderly fallers: When does the bilateral coordination of gait require attention? *Experimental Brain Research 177*(3), 336–346.
Yun Y, **Kim H-C**, **Shin SY**, **Lee J**, **Deshpande AD and Kim C** (2014) Statistical method for prediction of gait kinematics with Gaussian process regression. *Journal of Biomechanics 47*(1), 186–192.

Appendix A. Key standard GP equations

The reader should note the shorthand notation K_{XX} used to denote $k(X,X)$ throughout this appendix. Following the brief introduction to GPs in Section 2, the joint Gaussian distribution of the observed target values $\boldsymbol{y}$ and the function values at test locations $\boldsymbol{y}_\star$, under the prior is given by

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_\star \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(\boldsymbol{x}_\star) \end{bmatrix}, \begin{bmatrix} K_{XX} + \sigma_n^2 \mathbb{I} & K_{X\boldsymbol{x}_\star} \\ K_{\boldsymbol{x}_\star X} & K_{\boldsymbol{x}_\star \boldsymbol{x}_\star} + \sigma_n^2 \mathbb{I} \end{bmatrix} \right) \tag{15}$$

Here, X denotes a set of N, D-dimensional training inputs, where $X \in \mathbb{R}^{N \times D}$, whereas $\boldsymbol{y} \in \mathbb{R}^{N \times 1}$ is the corresponding set of N measured training outputs. Given the training inputs X and their corresponding outputs $\boldsymbol{y}$, the predictive distributions over $y_\star$ at new test input locations $\boldsymbol{x}_\star$ are given by

$$p(\boldsymbol{y}_\star | \boldsymbol{x}_\star, \boldsymbol{X}, \boldsymbol{y}) \sim \mathcal{N}(\mathbb{E}_{Standard}[\boldsymbol{y}_\star], \mathbb{V}_{Standard}[\boldsymbol{y}_\star]) \tag{16a}$$

$$\mathbb{E}_{Standard}[\boldsymbol{y}_\star] = m(\boldsymbol{x}_\star) + K_{\boldsymbol{x}_\star X}\left(K_{XX} + \sigma_n^2 \mathbb{I}\right)^{-1}(\boldsymbol{y} - m(\boldsymbol{x})) \tag{16b}$$

$$\mathbb{V}_{Standard}[\boldsymbol{y}_\star] = K_{\boldsymbol{x}_\star \boldsymbol{x}_\star} - K_{\boldsymbol{x}_\star X}\left(K_{XX} + \sigma_n^2 \mathbb{I}\right)^{-1} K_{X\boldsymbol{x}_\star} + \sigma_n^2 \mathbb{I} \tag{16c}$$

The kernel hyperparameters can be found via the following optimization of the negative log marginal likelihood:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\theta} - \log p(\boldsymbol{y} | \boldsymbol{x}, \boldsymbol{\theta}) \tag{17}$$

with

$$\begin{aligned} -\log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) &= -\log \mathcal{N}\left(\boldsymbol{y} | m(\boldsymbol{x}), K_{XX} + \sigma_n^2 \mathbb{I}\right) \\ &= \underbrace{\frac{N}{2}\log(2\pi)}_{\textit{constant}\text{ term}} + \underbrace{\frac{1}{2}\log|K_{XX} + \sigma_n^2 \mathbb{I}|}_{\textit{complexity}\text{ term}} + \underbrace{\frac{1}{2}\left((\boldsymbol{y} - m(\boldsymbol{x}))^T \left(K_{XX} + \sigma_n^2 \mathbb{I}\right)^{-1}(\boldsymbol{y} - m(\boldsymbol{x}))\right)}_{\textit{model}\text{ fit term}} \end{aligned} \tag{18}$$

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

The annotated terms in Equation 18 have readily available interpretations, and it can be clearly seen that there is a trade-off between model fit and model complexity. This property is known as the Bayesian Occam's Razor (Rasmussen and Ghahramani, 2001; Rasmussen and Williams, 2006). Thus, the hyperparameters of the kernel can be learnt and the GP is completely defined by Equations 15 and 16.

Appendix B. Key sparse GP equations

Following the introduction presented in Section 2.1, the key equations required for variational approximation methods are presented here. Briefly, variational inference methods derive a lower bound of the log-marginal likelihood using a distribution $q(f)$ over the entire infinite-dimensional function:

$$\log p(\boldsymbol{y}|\boldsymbol{\theta}) = \log \int p(\boldsymbol{y},f|\boldsymbol{\theta})df \geq \int q(f)\log\frac{p(\boldsymbol{y},f|\boldsymbol{\theta})}{q(f)}df = \mathbb{E}_{q(f)}\left[\log\frac{p(\boldsymbol{y},f|\boldsymbol{\theta})}{q(f)}\right] = F_{Sparse}(q,\boldsymbol{\theta}) \tag{19}$$

This lower bound F_{Sparse} can be expressed in terms of the KL divergence between the variational distribution and the true posterior:

$$F_{Sparse}(q,\boldsymbol{\theta}) = \mathbb{E}_q[\log p(\boldsymbol{y}|\boldsymbol{\theta})] - KL(q(f)\|p(f|\boldsymbol{y},\boldsymbol{\theta})) \tag{20}$$

Given the set of inducing points $\{Z,\boldsymbol{u}\}$, where Z contains the locations of the inducing points and $\boldsymbol{u}$ are the values of the latent functions at these points, the variational lower bound can be written explicitly as

$$\begin{aligned} F_{Sparse}(Z) = &-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log|Q_{XX}+\sigma_n^2\mathbb{I}| - \frac{1}{2}(\boldsymbol{y}-m(X))^T\left[Q_{XX}+\sigma_n^2\mathbb{I}\right]^{-1}(\boldsymbol{y}-m(X)) \\ &-\frac{1}{2}\sigma_n^{-2}tr(K_{XX}-Q_{XX}) \end{aligned} \tag{21}$$

where $tr(\cdot)$ is the trace operator and Q_{XX} is the approximate covariance matrix, defined as

$$Q_{XX} = K_{X\boldsymbol{u}}K_{\boldsymbol{uu}}^{-1}K_{\boldsymbol{u}X} \tag{22}$$

Here, the kernel functions, evaluated at the data points X, inducing input points Z, and between the data and inducing points, are represented by the kernel matrices K_{XX}, $K_{\boldsymbol{uu}}$ and $K_{X\boldsymbol{u}}$, respectively. Please note, that according to Titsias, 2009; Bui et al. (2017), the notation used here can be generalized, such that:

$$Q_{\boldsymbol{ab}} = K_{\boldsymbol{au}}K_{\boldsymbol{uu}}^{-1}K_{\boldsymbol{ub}} \tag{23}$$

The bound derived in Equation 21 can then be used for hyperparameter optimization. For the complete derivation of this bound, the reader is referred to Michalis (2009). Following optimization, predictions can be done in a comparable manner to the standard GP. Hence, the predictive distribution is given by (It should be noted that the explicit conditioning of the posteriors on the training data, test input, inducing points, and hyperparameters was dropped here for simplicity of notation.)

$$p(\boldsymbol{y}_\star|\boldsymbol{x}_\star,X,\boldsymbol{y},\boldsymbol{u}) = \mathcal{N}\left(\mathbb{E}_{Sparse}[\boldsymbol{y}_\star],\mathbb{V}_{Sparse}[\boldsymbol{y}_\star]\right) \tag{24a}$$

$$\mathbb{E}_{Sparse}[\boldsymbol{y}_\star] = Q_{\boldsymbol{x}_\star X}\left(Q_{XX}+\sigma_n^2\mathbb{I}\right)^{-1}\boldsymbol{y} \tag{24b}$$

$$\mathbb{V}_{Sparse}[\boldsymbol{y}_\star] = K_{\boldsymbol{x}_\star\boldsymbol{x}_\star} - Q_{\boldsymbol{x}_\star X}\left(Q_{XX}+\sigma_n^2\mathbb{I}\right)^{-1}Q_{X\boldsymbol{x}_\star} \tag{24c}$$

The main benefit of this sparse approximation is that the computational requirements are reduced from $\mathcal{O}(N^3)$ to $\mathcal{O}(NM^2)$ for M inducing points. Naively, it is not clear where the computational speed-up is found. However, note that the Woodbury inversion lemma can be applied as detailed below:

$$\begin{aligned} &\left(Q_{XX}+\sigma_n^2\mathbb{I}\right)^{-1} = \\ &\left(K_{Xu}K_{uu}^{-1}K_{uX}+\sigma_n^2\mathbb{I}\right)^{-1} = \\ &\sigma_n^{-2}\mathbb{I} - \sigma_n^{-2}\mathbb{I}\underbrace{\left(K_{uu}+K_{uX}\left(\sigma_n^{-2}\mathbb{I}\right)K_{Xu}\right)^{-1}}_{\mathbb{R}^{M\times M}}K_{uX}\sigma_n^{-2}\mathbb{I} \end{aligned} \tag{25}$$

Appendix C. Key heteroscedastic GP equations

The heteroscedastic lower bound is given by Equation 26, as follows:

$$F_{Heteroscedastic}(\boldsymbol{\mu},\Sigma) = \log\mathcal{N}\left(\boldsymbol{y}|\boldsymbol{0},K_f+R\right) - \frac{1}{4}tr(\Sigma) - KL(\mathcal{N}(\boldsymbol{h}|\boldsymbol{\mu},\Sigma)\|\mathcal{N}(\boldsymbol{h}|\mu_0\boldsymbol{1},K_h)) \tag{26}$$

To ensure clear notation, K_f and K_h are used here to denote the covariance matrices of the two GPs used to model $f(\boldsymbol{x})$ and $h(\boldsymbol{x})$, respectively. R is a diagonal matrix whose elements are $R_{ii} = \exp(\boldsymbol{\mu}_i - 0.5\Sigma_{ii})$. Moreover, $\boldsymbol{\mu}$ and Σ are the variational parameters to be determined, defined according to Equation 27, for some positive semidefinite diagonal matrix Λ (see Liu et al., 2018):

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

$$\boldsymbol{\mu} = K_h\left(\Lambda - \frac{1}{2}\mathbb{I}\right)\mathbf{1} + \mu_0\mathbf{1}, \tag{27a}$$

$$\Sigma^{-1} = K_h^{-1} + \Lambda \tag{27b}$$

It can be seen that this implementation requires the optimization of $N + N(N+1)/2$ free variational parameters. This is achieved through the reparametrization of $\boldsymbol{\mu}$ and Σ in terms of Λ (Lázaro-Gredilla and Titsias, 2011). As a result, the computational complexity is roughly twice that of the homoscedastic GP.

Another challenge that has emerged from the use of a heteroscedastic GP is the lack of a complete predictive distribution in a closed form. Fortunately, it is still possible to approximate the first two moments (that is the mean and variance) of the predictive distribution. These are expressed in the following equations:

$$\mathbb{E}_q[\boldsymbol{y}_\star] = \boldsymbol{a}_\star \tag{28a}$$

$$\mathbb{V}_q[\boldsymbol{y}_\star] = \boldsymbol{c}_\star{}^2 + \exp\left(\boldsymbol{\mu}_\star + \frac{1}{2}\sigma_\star{}^2\right) \tag{28b}$$

Here, the following notations have been used:

$$\boldsymbol{a}_\star = k_f(x_\star, X)\left(K_f + R\right)^{-1}\boldsymbol{y} \tag{29a}$$

$$\boldsymbol{c}_\star{}^2 = k_f(\boldsymbol{x}_\star, \boldsymbol{x}_\star) - k_f(\boldsymbol{x}_\star, X)\left(\left(K_f + R\right)^{-1}\right)k_f(X, \boldsymbol{x}_\star) \tag{29b}$$

$$\boldsymbol{\mu}_\star = k_h(\boldsymbol{x}_\star, X)\left(\Lambda - \frac{1}{2}\mathbb{I}\right)\mathbf{1} + \mu_0 \tag{29c}$$

$$\sigma_\star{}^2 = k_h(\boldsymbol{x}_\star, \boldsymbol{x}_\star) - k_h(\boldsymbol{x}_\star, X)\left(K_h + \Lambda^{-1}\right)^{-1}k_h(X, \boldsymbol{x}_\star) \tag{29d}$$

With these definitions, it is now possible to make predictions using a heteroscedastic GP. Here, it is assumed that the first two moments presented above are representative of the true underlying distribution.

Appendix D. Key variational sparse heteroscedastic GP equations

In this section, the covariance matrices are indexed by a superscript f or h, which denotes the function and corresponding hyperparameters under consideration. The subscripts denote which sets of points the covariance is computed between, with X being the full set of points and u being the set of inducing points for the corresponding function. Therefore, as an example, K_{Xu}^h is the covariance matrix between the training points and the inducing points for $h(\boldsymbol{x})$, given the hyperparameters of the kernel for the log-noise GP. Although there is a nontrivial amount of algebra to arrive at these equations, to learn the optimal set of hyperparameters for the VSHGP model, the problem reduces to maximizing the following lower bound (Liu et al., 2018), which is defined as

$$\begin{aligned} F_{Sparse\text{-}Heteroscedastic}(\boldsymbol{\mu}, \Sigma) &= \log\mathcal{N}\left(\boldsymbol{y}|\mathbf{0}, Q_{XX}^f + R_h\right) - 0.25Tr[\Sigma_h] \\ -0.5Tr\left[R_h^{-1}\left(K_{XX}^f - Q_{XX}^f\right)\right] &- KL\left(\mathcal{N}(\boldsymbol{h}|\boldsymbol{\mu}_{\boldsymbol{u}}, \Sigma_u)\|\mathcal{N}\left(\boldsymbol{h}|\mu_0, K_{uu}^h\right)\right) \end{aligned} \tag{30}$$

where the diagonal matrix $R_h \in \mathbb{R}^{N\times N}$ is defined as: $R_{ii} = \exp(\boldsymbol{\mu}_{\boldsymbol{hi}} - 0.5\Sigma_{hii})$, with the mean and variance.

$$\boldsymbol{\mu}_{\boldsymbol{h}} = \Omega_{Xu}^h(\boldsymbol{\mu}_{\boldsymbol{u}} - \mu_0\mathbf{1}) + \mu_0\mathbf{1} \tag{31a}$$

$$\Sigma_h = K_{XX}^h - Q_{XX}^h + \Omega_{Xu}^h\Sigma_u\left(\Omega_{Xu}^h\right)^T \tag{31b}$$

$$\boldsymbol{\mu}_{\boldsymbol{u}} = K_{uX}^h(\Lambda - 0.5\mathbb{I})\mathbf{1} + \mu_0\mathbf{1} \tag{31c}$$

$$\Sigma_u^{-1} = \left(K_{uu}^h\right)^{-1} + \left(\Omega_{Xu}^h\right)^T\Lambda\Omega_{Xu}^h \tag{31d}$$

Here, Λ is a positive semi-definite diagonal matrix, and Ω_{Xu}^h is defined according to Equation 32. For details regarding the full derivation, the reader is referred to Liu et al. (2018).

$$\Omega_{Xu}^h = K_{Xu}^h\left(K_{uu}^h\right)^{-1} \tag{32}$$

The sparse approximation for the two GPs employed to model the heteroscedastic noise introduces several additional hyperparameters associated with the inducing points used in $f(\boldsymbol{x})$ and $h(\boldsymbol{x})$. It should be noted that the number of inducing points does not necessarily have to be equal for both functions. As a result of the introduction of the two sets of inducing points, the number of hyperparameters has now increased to include the kernel hyperparameters for $k_f(\boldsymbol{x}, \boldsymbol{x}')$ and $k_h(\boldsymbol{x}, \boldsymbol{x}')$, the constant mean for the log noise variance, μ_0, the location of the m inducing points for $f(\boldsymbol{x})$, the location of the u inducing points for $h(\boldsymbol{x})$, as well as the n variational parameters composing the Λ diagonal matrix.

However, as with the non-sparse heteroscedastic GP, computing the predictive distribution $p(\boldsymbol{y}_\star|\boldsymbol{y}, \boldsymbol{x}_\star)$ at the test points $\boldsymbol{x}_\star$ requires marginalizing over the latent noise process, leading to an intractable integral. This marginalization implies that the predictive posterior is no longer Gaussian. In practice, following Lázaro-Gredilla and Titsias (2011) and Liu et al. (2018), we can

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

approximate the predictive distribution by matching its first two moments via Gauss–Hermite quadrature. This procedure yields an approximate Gaussian predictive distribution with mean $\mu_\star$ and variance $\sigma_\star^2$ as follows:

$$\boldsymbol{\mu}_\star = \boldsymbol{\mu}_\star^f,\ \sigma_\star^2 = {\sigma_\star}^{f^2} + e^{{\boldsymbol{\mu}_\star}^h + {\sigma_\star}^{h^2}/2} \tag{33}$$

where

$${\boldsymbol{\mu}_\star}^f = K_{\star u}^f K_R^{-1} K_{uX}^f R_h^{-1} y, \tag{34a}$$

$${\sigma_\star}^{f^2} = K_{\star\star}^f - K_{\star u}^f \left(K_{uu}^f\right)^{-1} K_{u\star}^f + K_{\star u}^f K_R^{-1} k_{u\star}^f, \tag{34b}$$

$${\sigma_\star}^{h^2} = K_{\star\star}^h - K_{\star u}^h \left(K_{uu}^h\right)^{-1} K_{u\star}^h + K_{\star u}^h K_\Lambda^{-1} K_{u\star}^h, \tag{34c}$$

$$K_R = K_{uX}^f R_h^{-1} K_{Xu}^f + K_{uu}^f \tag{34d}$$

$$K_\Lambda = K_{uX}^h \Lambda^{-1} K_{Xu}^h + K_{uu}^h \tag{34e}$$

Here, it is important to note that the correction term, $K_{\star u}^f K_R^{-1} K_{u\star}^f$ in Equation 34b contains the heteroscedasticity information from the noise term R_h.

Appendix E. Group and individual model comparisons using the KL divergence

Figure E1 demonstrates the asymmetric nature of the Kullback–Leibler (KL) divergence when comparing group-level differences across the gait cycle. The left panel shows the divergence pattern when using the healthy control group as the reference distribution, while the right panel shows the same comparison with the MS group as the reference. The distinct patterns between these two panels highlight a key limitation of KL divergence: its asymmetric property leads to different interpretations depending on the choice of reference distribution. In addition to this, it is clear that the KL divergence is disproportionately sensitive to small values in the reference distribution rather than capturing the absolute magnitude of differences. As such, both visualizations show notable dissimilarities in the first 15%, the range of 35 to 55%, and the range of 60 to 90% of the gait cycle, but with different magnitudes and temporal distributions, emphasizing why a symmetric metric like the Maximum Mean Discrepancy (MMD) might be preferable for clinical interpretations (see Section 3.3). This same analysis is extended to individual-level comparisons against the control group, as shown in Figure E2.

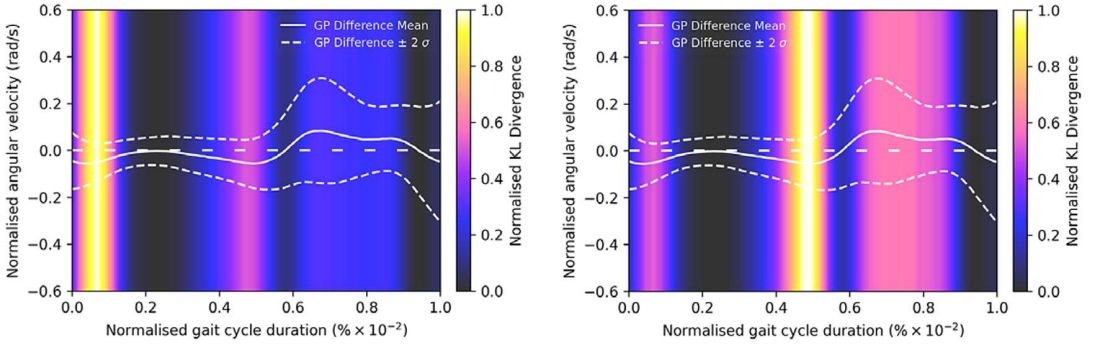

Figure E1. *Comparison of GP posterior distributions using KL divergence. (Left) KL divergence computed using the healthy control (HC) group GP posterior as the reference distribution. (Right) KL divergence computed using the MS group GP posterior as the reference distribution. The solid white line represents the mean difference between the two GPs, while the dashed lines indicate the* $\pm 2\sigma$ *uncertainty bounds of the difference.*

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

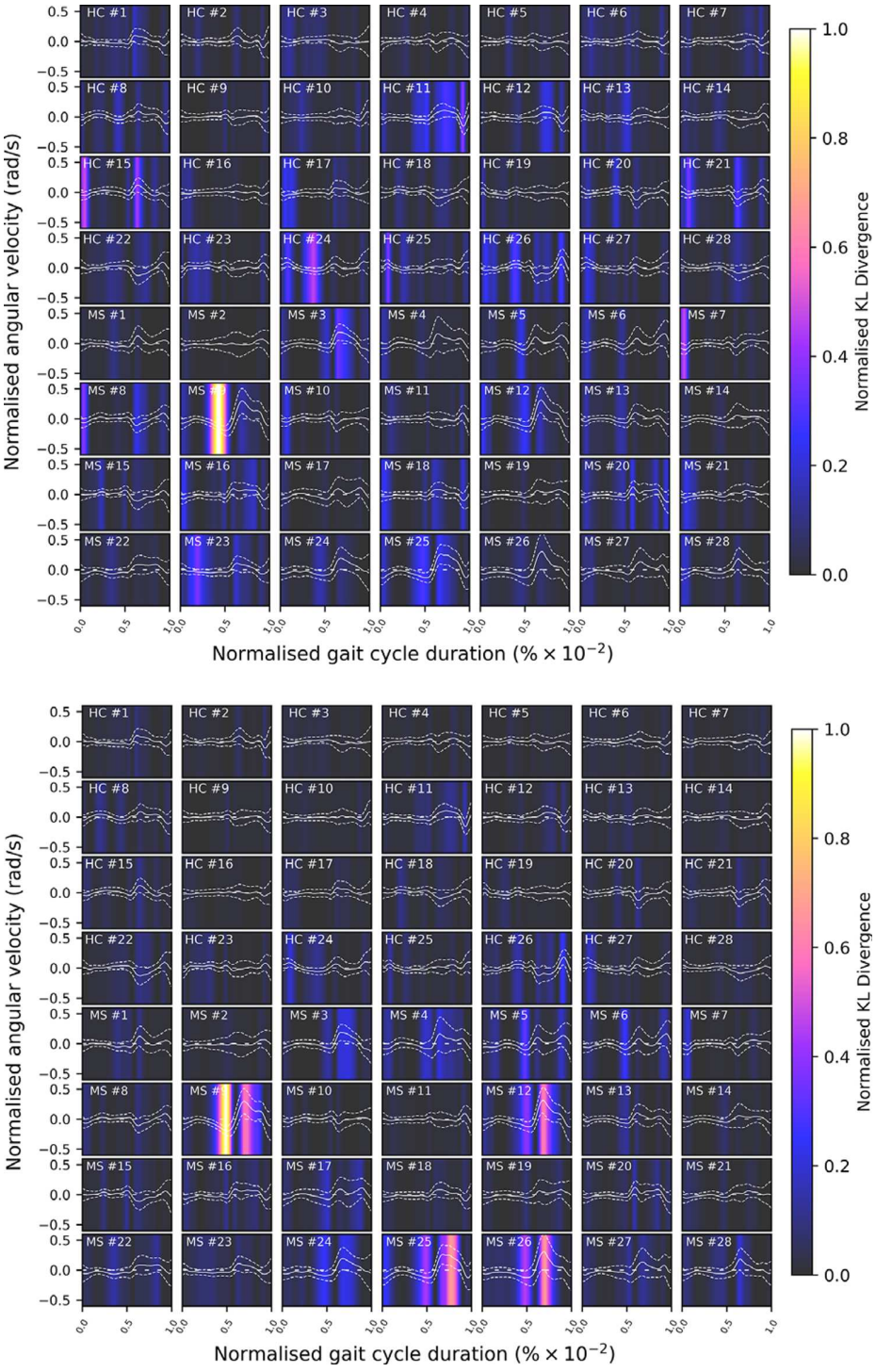

Figure E2. *Comparison of individual differences using KL divergence. (Top) KL divergence computed using the healthy control (HC) group GP posterior as the reference distribution. (Bottom) KL divergence computed using individual-specific posteriors as reference distributions. The first four rows represent HC individuals, while the last four rows represent people with MS (PwMS). Higher values (yellow) indicate greater divergence between distributions.*

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

Appendix F. Individual models: Performance metrics

The individual model (aggregating contralateral limb data) performance metrics are presented in Table F1.

Table F1. *Performance metrics of individual models aggregating contralateral limb data*

Healthy controls					PwMS					
	NMSE (%)		MSLL				NMSE (%)		MSLL	
Subject no.	Train	Test	Train	Test	Subject no.	EDSS	Train	Test	Train	Test
1	0.841	0.885	−4.177	−4.183	1	4.5	2.838	2.821	−3.870	−3.838
2	1.303	1.277	−4.025	−4.023	2	4.5	3.914	3.925	−3.554	−3.528
3	1.012	0.972	−4.153	−4.134	3	4	1.381	1.347	−4.193	−4.208
4	1.260	1.286	−4.086	−4.092	4	4	8.810	9.162	−3.609	−3.584
5	0.591	0.619	−4.135	−4.109	5	5	6.227	6.344	−3.736	−3.773
6	0.562	0.555	−4.321	−4.320	6	5	4.573	4.589	−3.736	−3.711
7	0.557	0.563	−4.245	−4.231	7	4.5	1.434	1.413	−4.167	−4.176
8	1.014	1.046	−4.108	−4.099	8	2	0.877	0.897	−4.122	−4.124
9	0.742	0.735	−4.087	−4.091	9	3.5	7.875	7.907	−4.677	−4.690
10	0.588	0.598	−4.292	−4.278	10	3.5	1.000	1.032	−4.027	−4.018
11	0.634	0.654	−4.474	−4.456	11	2	0.736	0.738	−4.245	−4.238
12	0.793	0.801	−4.218	−4.207	12	5	6.190	6.275	−4.127	−4.132
13	0.626	0.646	−4.382	−4.366	13	4.5	2.062	2.192	−4.068	−4.030
14	0.992	0.972	−4.109	−4.078	14	2.5	2.899	2.942	−3.887	−3.884
15	0.593	0.586	−4.194	−4.193	15	2	1.434	1.429	−3.878	−3.852
16	0.810	0.813	−4.114	−4.127	16	1.5	1.013	1.004	−4.173	−4.187
17	0.778	0.774	−4.175	−4.173	17	3.5	6.648	6.310	−3.549	−3.570
18	1.108	1.113	−3.913	−3.925	18	2	0.620	0.649	−4.237	−4.211
19	0.800	0.790	−4.029	−4.049	19	2	4.265	4.154	−3.772	−3.801
20	0.927	0.929	−4.087	−4.043	20	2	0.966	0.960	−4.315	−4.293
21	0.575	0.583	−4.366	−4.381	21	3.5	1.320	1.335	−4.191	−4.203
22	1.658	1.662	−3.880	−3.882	22	2	2.933	2.919	−3.998	−3.997
23	0.952	0.973	−4.172	−4.180	23	2	0.729	0.763	−4.307	−4.291
24	1.054	1.093	−4.219	−4.204	24	3.5	5.856	5.974	−3.831	−3.775
25	0.703	0.696	−4.424	−4.433	25	3.5	5.321	5.434	−4.473	−4.438
26	0.973	1.007	−4.055	−4.013	26	4	11.427	11.012	−3.571	−3.553
27	1.094	1.094	−4.098	−4.092	27	4	6.722	6.684	−3.664	−3.663
28	0.637	0.634	−4.073	−4.064	28	4	2.792	2.787	−3.807	−3.823

https://doi.org/10.1017/dce.2025.10009 Published online by Cambridge University Press

Appendix G. Wasserstein asymmetry

Figure G1 displays the distinctive locations across the entire gait cycle where asymmetry is present in the gait patterns obtained from the individual limb models, that is left versus right. Here, each individual line corresponds to a unique individual.

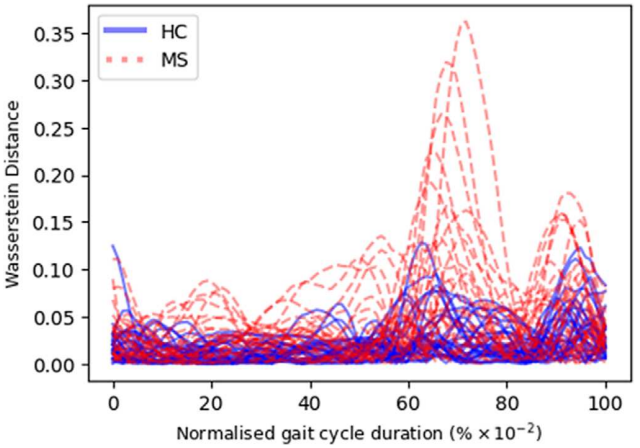

Figure G1. *Wasserstein distance computed between left and right limb models.*

Cite this article: Stihi A, Mazzà C, Cross E and Rogers TJ (2025). Hierarchical Gaussian processes for characterizing gait variability in multiple sclerosis. *Data-Centric Engineering*, 6, e36. doi:10.1017/dce.2025.10009