*Article*

# Reducing the Complexity of Musculoskeletal Models using Gaussian Process Emulators

**Ivan Benemerito [1,†,*], Erica Montefiori [1,†], Alberto Marzo [1] and Claudia Mazzà [1]**

[1] Department of Mechanical Engineering & INSIGNEO Institute for *in silico* Medicine, The University of Sheffield, UK
* Correspondence: i.benemerito@sheffield.ac.uk
† These authors contributed equally to the study

**Abstract:** Musculoskeletal models (MSKMs) are used to estimate the muscle and joint forces involved in human locomotion, often associated with the onset of degenerative musculoskeletal pathologies (e.g. osteoarthritis). Subject-specific MSKMs offer more accurate predictions than their scaled-generic counterparts. This accuracy is achieved through time-consuming personalisation of models and manual tuning procedures that suffers from potential repeatability errors, hence limiting the wider application of this modelling approach. In this work we have developed a methodology for identifying and ranking the muscles that are more important to the determination of the joint forces, thus producing reduced but still accurate representation of the musculoskeletal system in shorter timeframes. The methodology hinges on Sobol's sensitivity analysis (SSA) for ranking the muscle importance. The thousands of data points required for SSA are generated using Gaussian Process emulators, a Bayesian technique to infer the input-output relationship between nonlinear models from a limited number of observations. Results show that there is a pool of muscles whose personalisation has little effects on the model predictions. Furthermore, joint forces in subject generic and subject generic models are influenced by different set of muscles, suggesting the existence of a model specific component of the sensitivity analysis.

**Keywords:** statistical modelling; statistical emulators; sensitivity analysis; Gaussian Process; Sobol; musculoskeletal model

## 1. Introduction

Musculoskeletal models (MSKMs) are a commonly adopted solution to estimate biomechanical parameters otherwise not directly measurable, to predict the outcome of interventions, to inform rehabilitation planning, or to test complex scientific hypothesis within clinical context [1-3]. MSKMs can be divided into two main categories: generic, and subject specific. Generic models are constructed by scaling a reference model using anthropometric measures [1, 4], while subject-specific models rely on medical images and their segmentation for the personalisation of anatomical features and input parameters [5-8]. They are known to provide accurate estimates of specific biomechanical quantities for an individual [8, 9]. Amongst these quantities, individual muscle forces and joint contact forces (JCFs) are of particular interest in the assessment of joint loading and have been

 doi:10.20944/preprints202211.0428.v1



linked to the onset and progression of degenerative diseases of the musculoskeletal system (i.e., juvenile idiopathic arthritis [3] and osteoarthritis [10, 11].

The creation of a subject-specific model is a time-consuming and operator-dependent process, which limits the applicability of imaging-based MSK modelling protocols. Overcoming this limitations would widen their usability across the biomechanical community, and ultimately in the clinical practice. Sensitivity studies can identify the parameters that influence models' output and inform strategies that guide limit model personalisation and pre-processing time to the minimum while still guaranteeing the accuracy of the predictions [12].

Several studies reported differences in model outputs when using MRI-based or generic anatomy [10, 13, 14], different joint types and degrees of freedom [15-17], or different muscle parameters and geometry [6, 7, 12]. With regards to muscle parameters in particular, the use of models with different characteristics may lead to contrasting findings or results that are not directly comparable and hence raises the question whether a more comprehensive study should include different types of models and participants to test model- and subject-dependency of model sensitivity. Navacchia and co-workers [18] conducted a sensitivity study combining several layers of input uncertainties and compared results across three participants to test subject-dependency. However, they based the analysis on a generic model and used a Monte Carlo approach relying on a limited number of simulations. In fact, to test the sensitivity of a model with $N$ input parameters $\mathcal{O}(1000*N)$ model evaluations are typically required [19], a task that can prove prohibitive even for relatively small models. Statistical emulators such as Gaussian Process (GPs) can emulate the input-output relationship of complex nonlinear systems using only a limited number of simulator runs as training points, in the order of O(N) [20]. The lower computational effort required thus makes them ideal for generating the points needed to assess the influence of complex MSKM model parameters on the output of interest, and to represent virtual populations with a much reduced computational effort, and while preserving output accuracy [21]. GPs have been adopted in several fields of research, spanning from prediction of scalar values in cardiac electrophysiology [22], to time series in finance [23] and weather forecast [24]. Applications in the musculoskeletal domain are still rare and focus on the estimation of muscle activation from surface electromyography signals [25] or on learning tasks in robotics [26].

Existing sensitivity studies conducted on MSKMs reported that tendon slack length and optimal fibre length are amongst the most important parameters for characterising muscle behaviour and influencing JCFs, followed by maximal isometric force ($F_{max}$) and others [12, 18, 27]. However, tendon slack length can only be characterised through cadaveric studies, and optimal fibre length is still not easily measurable unless using *ad hoc* imaging protocols [8], which limits the possibility of performing extensive and quantitative evaluation of their role. On the contrary, $F_{max}$ of a muscle, which is the maximal force that a muscle can express, can be easily estimated from MRI as a function of muscle cross-sectional area [28], measured through muscle segmentation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 November 2022 doi:10.20944/preprints202211.0428.v1



In this work we hypothesise that, given a distribution of muscle $F_{max}$, it is possible to use GPs for emulating the waveforms of JCFs resulting from generic and subject-specific MSKMs, and that they can be used to efficiently conduct comprehensive sensitivity analysis (SA) and investigate potential model-dependency of SA. Furthermore, we make the hypothesis that SA can inform model reduction strategies leading to minimisation of modelling time and costs. We aim to rank the muscles of the lower limb according to their importance in contribution to the determination of hip, knee, and ankle JCFs during a walking task. Results show that GPs can predict JCFs with high accuracy, and that through SA it is possible to identify the muscles that contribute more to the determination of the JCFs. Reduced models developed according to the proposed model reduction strategy present low errors when compared to their fully personalised counterparts.

## 2. Materials and Methods

This study, whose workflow is depicted in Figure 1, has three main parts: the mechanistic model development part (data collection, generation of MSKMs and dynamic simulation), the statistical emulation part (emulation and sensitivity analysis of MSKMs), and the model reduction part.

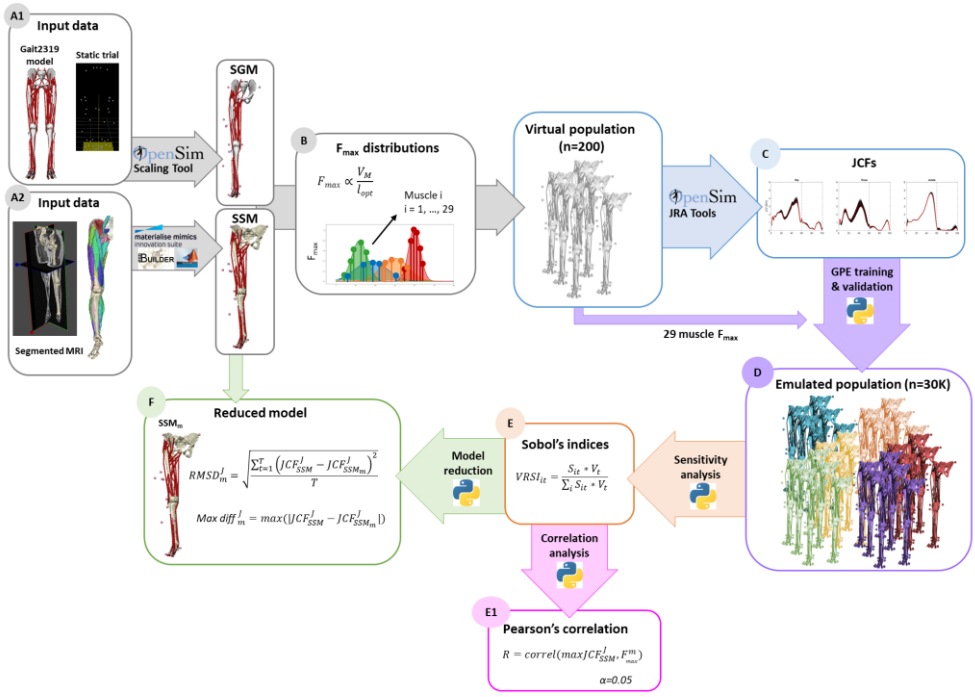


***Figure 1.*** *Workflow of the study. From top left clockwise, (A1) literature Gait2392 model is scaled in OpenSim using a static standing trial to obtain a scaled-generic model (SGM ) while (A2) MRI segmentation is used to generate a subject-specific model; (B) literature $F_{max}$ values from 29 lower-limb muscles are used to estimate distributions of $F_{max}$ across a virtual population (n=200); (C) joint contact forces (JCF) are estimated for hip, knee and ankle and input to the Gaussian Process emulator (GP) together with associated $F_{max}$ values for training and validation; (D) JCF waveforms are emulated (n=30K) to conduct (E) a Sobol's sensitivity analysis, whose results are then used to inform model reduction strategy; (E1) Pearson's correlation between JCF and $F_{max}$ of most*

 doi:10.20944/preprints202211.0428.v1



*influential muscles is assessed; (F) reduced model is compared, through RMSD and absolute maximum difference, to nominal fully-personalised model based.*

### 2.1 Mechanistic modelling

#### 2.1.1 Input data and baseline musculoskeletal models

This study used legacy data collected from post-menopausal women (UK EPSRC Multisim Study approved by the Health Research Authority of East of England and Cambridgeshire and Hertfordshire Research Ethics Committee, reference 16/EE/0049) [29, 30]. We selected a representative participant (74.6 y, 56.8 kg, 163.5 cm, BMI = 21.2) for whom 3D gait analysis data (marker trajectories and ground reaction forces) and MRI were available (full details of experimental data are described in [30]. Two different baseline monolateral MSKMs were built, one scaled-generic and one subject-specific. Each model included four body segments (pelvis, femur, tibia, foot) articulated by an ideal ball-and-socket joint for the hip, and two ideal hinges, one for the knee and one for the ankle and 43 lower-limb muscles. However, they differed for how the joint axes were defined and for the values of $F_{max}$ assigned to the individual muscles. The details of the two models are as follow:

- **Scaled-generic model (SGM).** Marker-based scaling of a literature model (Gait2392, [1]) allowed to obtain a scaled generic model of the participant's limb. Scaling was performed according to best practice recommendation [31] using a static standing trial and the OpenSim Scaling Tool [32]. Muscle $F_{max}$ was linearly scaled [32] from Gait2392 default values based on participant's body mass.
- **Subject-specific model (SSM).** MRI segmentation enabled the generation of a subject-specific model with personalised bone geometries and segment inertia [7] and joint axes determined via morphological fitting to the articular surface of the segmented bone geometries. The same set of muscles included in SGM was included in SSM but their origin, insertion and via points were personalised based on the MRI. In SSM, muscle length parameters were linearly scaled from Gait2392 values in order to maintain their ratio to musculotendon length. $F_{max}$ of 29 lower-limb muscles was personalised using MRI-segmented muscle volume available from the online free repository associated to a study by Montefiori and colleagues [30], comprising of eleven older women (including the participant used in this study) enrolled as part of the above-mentioned Multisim project [cite repos]. $F_{max}$ of the remaining 14 muscles (not available in the above-mentioned repository as deemed as not-repeatably measurable) were linearly scaled from Gait2392 values based on body mass of the participant.

#### 2.1.2 Virtual population

Four hundred models (200 variations of SGM and 200 variations of SSM) were generated to build a virtual population of individuals. Mean and standard deviation values of the $F_{max}$ of the considered muscles were calculated from the cohort values reported by Montefiori et al. 2020 [30] and used to generate normal distributions of $F_{max}$ that were considered representative of a virtual population of older women. Independent random sampling of each muscle $F_{max}$ distribution allowed to determine 200 sets of 29 muscle $F_{max}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 November 2022 doi:10.20944/preprints202211.0428.v1



that were then used to characterise muscle properties of each model. The number of sampling points was based on a convergence study and was chosen to ensure less than 10% error in the normalised overlap of the resulting JCF curve bands.

*2.1.3 Dynamic simulations and data analysis*

Hip, knee, and ankle joint angles and moments were computed from the baseline models (SGM and SSM) using the OpenSim 3.3 [32] Inverse Kinematics and Inverse Dynamics Tools relying on the MATLAB API (v9.1, R2021b, Mathworks, USA). OpenSim recommended good practices [31] were followed. Two-hundred runs of Static Optimisation (where the sum of muscle activations squared was minimised) and Joint Reaction Analysis [33] enabled the estimate of individual muscle forces and associated JCFs norms for each virtual case. Ideal moment generators (reserve actuators), providing joint torque when muscle forces could not balance the external moments, were included for each degree of freedom, but made unfavourable to recruit by assigning them a unitary maximum force. Range of JCFs obtained with the 200 simulations were quantified and maximum percentage variation with respect to peak baseline values were calculated.

*2.2 Statistical modelling*

*2.2.1 Gaussian Process Emulator*

The 200 sets of 29 $F_{max}$ served as input for a GP (built using the GPy Python3 library [34]), which was trained to output the corresponding hip, knee, and ankle JCFs. To avoid ill conditioning of the emulator, both inputs and outputs were normalised prior to training. One emulator (zero mean, kernel: squared exponential plus Matern52) was trained for each joint of both SGM and SSM. Following standard practice in the field of GP, the performance of the emulator was assessed on the validation dataset through the mean average percentage error (MAPE), defined as:

$$MAPE = \frac{100\%}{N_v}\frac{1}{\bar{y_s}}\sum_{n=1}^{N_v}|y_s^n - y_e^n|$$

where $y_s^n$ and $y_e^n$ are the $n$-th run of the simulator and emulator respectively, $\bar{y_s}$ is the time average of the simulator output, and $N_v$ is the number of points in the validation dataset [35, 36]. Following a convergence study, the size of the training dataset which provided a low emulation error (MAPE<3%) while identified the computational cost was chosen to be 50 points.

*2.3 Model reduction*

*2.3.1 Sobol's sensitivity analysis*

In order to assess the model-dependency of SA, a global Sobol's sensitivity analysis [37] was performed independently on SGM and SSM. This allowed to evaluate the contribution of individual muscle $F_{max}$ variations to overall variations in the output JCFs. SA decomposes the output variance and ranks the contribution of individual inputs by mean of the Sobol's indices, real numbers ranging from 0 to 1, with 1 signifying that the entire

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 November 2022 doi:10.20944/preprints202211.0428.v1



variability of the output can be ascribed to the variation of a single input. The Saltelli algorithm [19] was used to sample the input parameter space and generate 30720 virtual subjects. For each virtual subject, the values of each $F_{max}$ were sampled from a normal distribution with mean and standard deviation derived from the measured distributions. The number of virtual subjects was chosen to guarantee convergence of the Sobol's algorithm, which was implemented through the SALib Python library [38]. The trained GPs were used to predict the hip, knee, and ankle JCF waveforms for each virtual subject.

Sobol's indices ascribe fractions of the output variance to variations of individual outputs, but do not account for the size of the output variance. To account for this a new metric, VRSI (variance renormalised Sobol's indices), was defined in order to normalise the Sobol's indices based on the size of the output variance:

$$VRSI_{it} = \frac{S_{it} * V_t}{\sum_i S_{it} * V_t}$$

where $S_{it}$ is the Sobol index of input *i* with the JCF at time *t*, and $V_t$ is the variance of the JCF at time t. By means of this normalisation process, we deemed as less important those inputs that showed high Sobol's index in a region of the gait cycle where the output variance was small.

A further analysis was conducted on the SSM data using the emulated dataset and the VRSI values. Pearson's Product-Moment correlation (alpha = 0.05) was calculated between the input $F_{max}$ and the peak values of JCF only for those muscles presenting a VRSI above 0.1 at the peaks. Significant correlations, with either moderate ($|R|>0.5$) or strong ($|R|>0.7$) correlation coefficient, were presented and discussed.

*2.3.2 Muscle ranking*

A reduced MSKM model was defined as a model having a reduced number of personalised muscle $F_{max}$, selected according to a ranking strategy based on the muscles contribution to the determination of total JCF. Similarly to previous SA studies [35, 36], a threshold of VRSI = 0.1 was set to identify the muscles that required personalisation. For each of the 100 frames of the gait cycle, the muscles that showed VRSI ≥ 0.1 for at least one of the JCFs were identified and ranked based on the number of time frames where they were deemed as influential. This allowed to reduce the number of muscles (m = 0 to 29) for which $F_{max}$ needs to be personalised to those with the highest ranking. Variations of SSM (referred to as $SSM_m$) were generated by decreasing m according to the above strategy: $SSM_{29}$ was equivalent to SSM (all 29 $F_{max}$ were personalised) while $SSM_0$ had $F_{max}$ linearly scaled from literature values (model Gait2392) [1].

JCFs for SSM and $SSM_m$ of the virtual subjects were then predicted with the emulator. To assess the extent to which reduced models could be used as a surrogate for SSM, we computed the maximum absolute difference (Δmax) as well as the root mean square deviation (RMSD) between the JCF of joint *J* estimated with $SSM_m$ and SSM over time *T* according to the following equations, respectively:

$$\Delta\max_m^J = \max(\left|JCF_{SSM}^J - JCF_{SSM_m}^J\right|)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 November 2022 doi:10.20944/preprints202211.0428.v1



$$RMSD_m^J = \sqrt{\frac{\sum_{t=1}^{T}(JCF_{SSM}^J - JCF_{SSM_m}^J)^2}{T}} \quad (228)$$

## 3. Results

### *3.1 Mechanistic simulations*

The simulation of 200 variations of the baseline SGM and SSM led to the JCF curves plotted in Figure 2. JCFs from the baseline models always laid within the range obtained from the 200 variations of each model. Sampling of $F_{max}$ led to variations of JCFs up to 0.8 BW, 1.1 BW, and 1.7 BW with SGM and up to 1.4 BW, 1.2 BW, and 2.1 BW with SSM for hip, knee, and ankle, respectively. These always occurred during late stance peak and corresponded to 21%, 33%, and 20% (SGM) and to 38%, 104%, and 59% (SSM) of baseline value for hip, knee, and ankle, respectively.

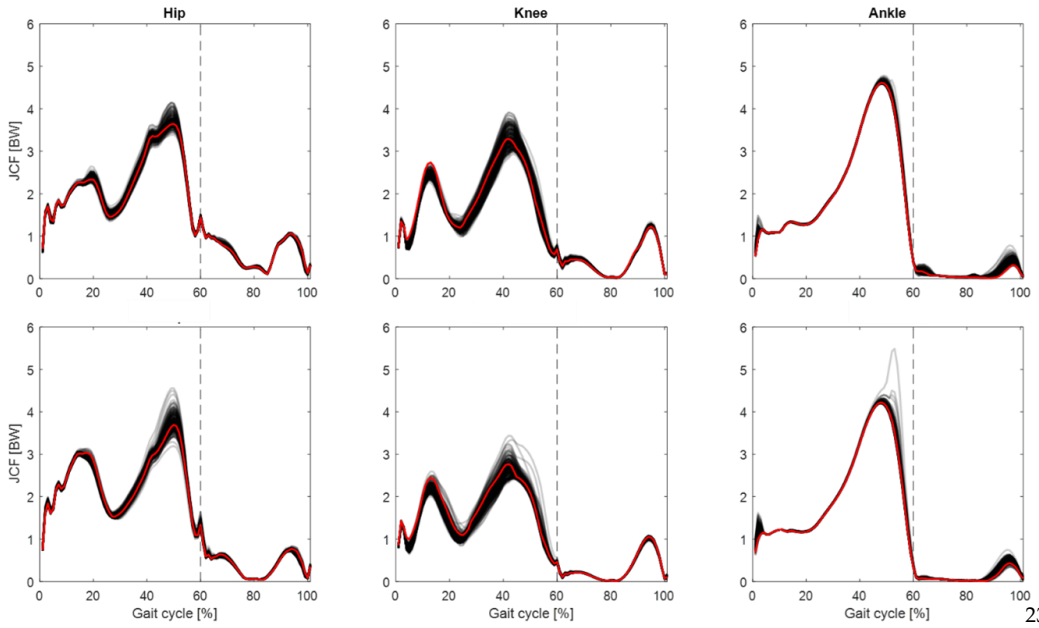


***Figure 2.*** *Two hundred JCF curves obtained by sampling $F_{max}$ distribution (black lines) and baseline curve (red line) obtained with SGM (top row) and SSM (bottom row); dashed vertical line indicates toe-off.*

### *3.2 Sobol's sensitivity analysis*

Overall, the sets of muscles whose $F_{max}$ had a large contribution to the determination of JCFs were different between SSM and SGM, as shown by the VRSIs of hip, knee, and ankle (heatmap in Figure 3). SGM had more muscles with VRSI ≥ 0.1 compared to SSM. Gracilis appeared to be influential for most of the stance on SGM's hip JCF (VRSI>=0.1 for 36% of the entire gait cycle, with peak value 0.6), with non-negligible contributions also to the knee JCF (VRSI>=0.1 for 16% of the gait cycle, with peak value 0.4). On the contrary, it was not influential in the determination of SSM's JCFs. Similar behaviour was observed also for Tensor Fasciae Latae, Adductor Magnus 1, Sartorius, and Semimembranosus. SSM's hip JCF was mainly influenced by Gluteus Medius 1 and Gastrocnemius Medialis, while the determination of knee JCF was mostly influenced by Gluteus Medius 1,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 November 2022 doi:10.20944/preprints202211.0428.v1



Gastrocnemius Medialis and Rectus Femoris. VRSIs tended to be low in the ankle joint because of the small variance of the input JCF across the virtual population. In the peak region of ankle JCF, the muscles showing the largest influence were Tibialis Posterior and Vastus Lateralis in SGM, and Soleus and Rectus Femoris in SSM.

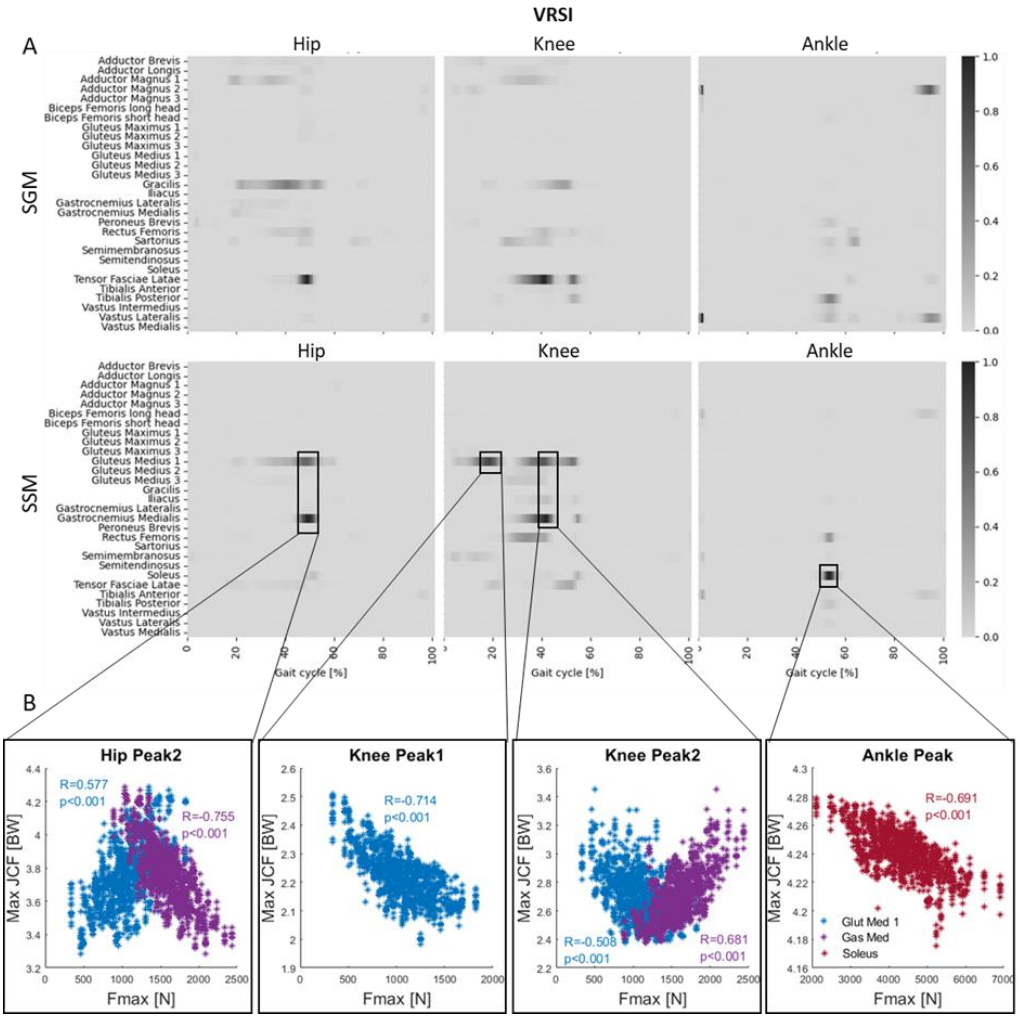


***Figure 3.*** *(A) VRSI heatmaps for JCFs of hip (left column), knee (central column) and ankle (right column). Results from generic model SGM are shown in top row, those from personalised SSM are in the bottom one. The x-axis represents the percentage of gait cycle, while the muscles are listed on the y-axis. Darker colours signify higher influence of the muscle on a specific time point. (B) Significant moderate (|R|>0.5) or strong (|R|>0.7) correlations between muscle $F_{max}$ and joint peak forces. For hip and knee, Peak 1 refers to loading acceptance phase of gait and Peak 2 refers to push off phase, while ankle Peak corresponds to push off phase.*

Correlation analysis confirmed the results of the sensitivity analysis (scatter plots in Figure 3). The $F_{max}$ of Gluteus Medius 1 had a moderate positive correlation (R=0.577) with Peak 1 of hip force, and a negative correlation with Peak 1 and Peak 2 of knee force (R=-0.714; R=-0.508, strong and moderate, respectively). The $F_{max}$ of Gastrocnemius medialis had a strong negative correlation (R=-0.755) with Peak 1 of hip force and had a moderate positive correlation (R=0.681) with Peak 2 of knee force. The $F_{max}$ of Soleus had a moderate

negative correlation (R=-0.691) with ankle peak force. All reported correlations had a p value below 0.001.

*3.3 Model reduction*

Of the 29 muscles in SSM, Gluteus Medius 1 was the most influential (Figure 4): with VRSI ≥ 0.1 for more than 40% of the gait cycle this muscle was the first to be personalised. Gastrocnemius Medialis, Rectus Femoris and Tensor Fasciae Latae followed, being influential for 19%, 16% and 10% of gait cycle, respectively. Semimembranosus and Iliacus were influential for 9% of the gait cycle, Soleus for 8%, Iliacus for 6%, while Tibialis Anterior, Biceps Femoris Longus, Tibialis Posterior were influential for less than 5% of the cycle. The remaining nineteen muscles never reached the threshold of VRSI = 0.1 and thus were not deemed as influential, and therefore their $F_{max}$ was never personalised.

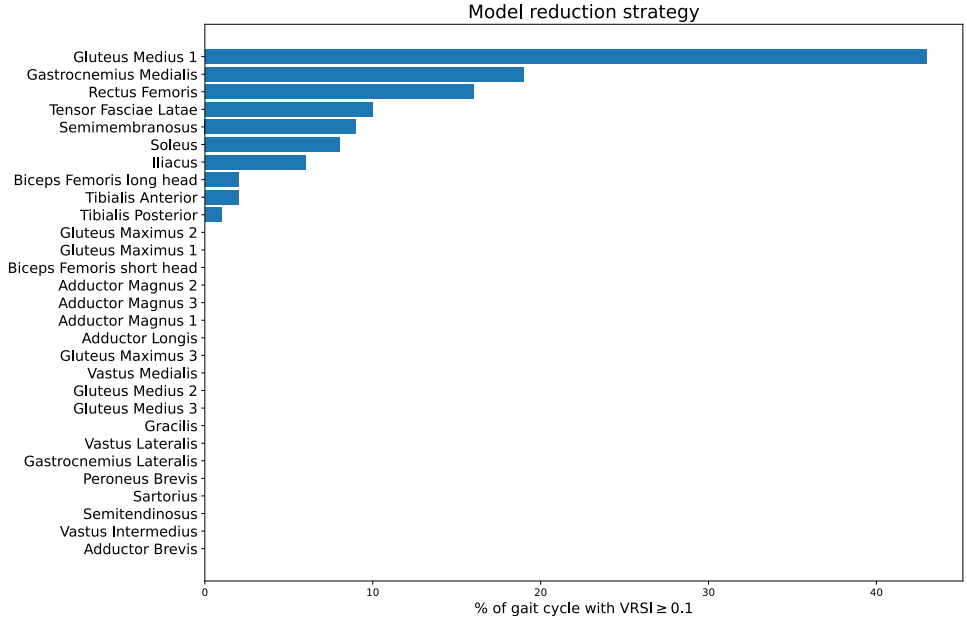


***Figure 4.*** *Percentage of gait cycle where VRSI ≥ 0.1 for the muscles included in the analysis. The chosen model reduction strategy prioritises the personalisation of muscles at the top of the plot, where VRSI is above 0.1 for a larger percentage of gait cycle.*

RMSDs between JCFs estimated with SSM and $SSM_m$ (reported as a function of $m$ in Figure 5) were overall very small (below 0.1 BW). The highest RMSD values were found at the knee joint when low degrees of personalisation were employed (RMSD = 0.08 BW, against 0.05 BW for both hip and ankle), and decreased when more muscles were personalised. The same trend was observed for hip and ankle. Eventually, the deviation between the reduced models and SSM reached zero when all the muscles in $SSM_m$ were personalised.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 November 2022 doi:10.20944/preprints202211.0428.v1



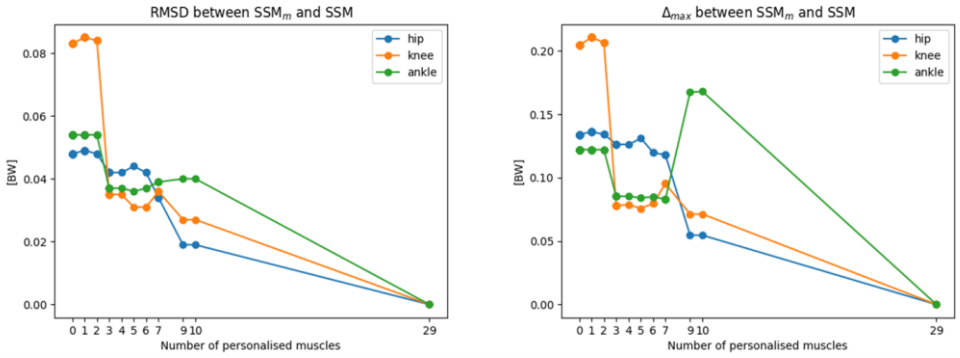


*Figure 5. RMSD (left) and Δmax (right) between fully-personalised model SSM and reduced models SSM$_m$, as a function of the number of personalised muscles.*

Similarly to RMSD, Δmax between SSM and SSM$_m$'s JCFs were always small (up to 0.2 BW). A value of 0.2 BW of difference was found at the knee when no muscles were personalised. This dropped to 0.07 BW with the personalisation of 9 muscles. Similarly, hip Δmax dropped from 0.14 BW to 0.06. At the ankle joint the behaviour was more complex, with an initial reduction in Δmax followed by a steep increase first, and then a final decrease.

## 4. Discussion

This study proposes a methodology for comprehensive sensitivity analyses of MSKMs' outputs to changing model parameters such as bone geometry, joint type and degrees of freedom, muscle parameters and geometry at the same time. This can offer a clearer insight into the model response and which input parameters must be personalised. To test the methodology, the feasibility of using GPs to emulate the JCF prediction of lower-limb MSK models was investigated, under the hypothesis that a trained GP could be used to identify the muscles whose $F_{max}$ is most influential to the estimate of JCFs. Notably, the methodology presented in this study has the potential to be applied to any input parameters given their statistical distributions.

In the first part of the study, we obtained JCF curves from 200 sets of OpenSim simulations (both for scaled-generic and subject-specific models) while varying $F_{max}$ of 29 limb muscles by 21 ± 6%, based on literature measurements [30]. Resulting simulated curves were overall similar in shape to literature reference curves obtained from MSKMs [7, 17, 18]. When comparing scaled-generic and subject-specific models, some qualitative differences were found between the scaled-generic and the subject-specific models. SGM had higher late stance peaks (for all joints) and higher knee early stance peak, while SSM had higher hip early stance peak. These were likely associated to differences in the definition of the joint axes, leading to different joint kinematics and associated actuation strategies as already hypothesised in previous literature [13, 16, 39]. As a consequence of changing $F_{max}$, we found variations up to 104% of baseline values (knee joint of SSM). The peak JCF of hip and knee corresponding to the push off phase of gait was particularly affected by the variations of $F_{max}$, in line with previous findings [7, 18].

 doi:10.20944/preprints202211.0428.v1



The sensitivity analysis was performed on data points that were artificially generated by the emulator, whose validation is of paramount importance to the credibility of the SA and the consequent model reduction strategy. In our case the emulator proved able to capture the entire variability in JCFs of the validation dataset using only 50 training points. The excellent behaviour on a validation dataset three times larger than the training one, clearly shows that the emulator training has been effective, and the risk of overfitting has been averted [40]. When applied to SSM and SGM this sensitivity analysis identified two distinct sets of muscles whose $F_{max}$ contribute to the determination of total JCFs. This was likely due to the above-mentioned differences in joint definition and in the values assigned to $F_{max}$, leading to different actuation strategies. This also confirmed the existence of a component of model-dependency in SA outcomes, as hypothesised in previous literature [13, 18], and showed that different sets of muscle $F_{max}$ are influential in determining total JCFs when using generic-scaled or subject-specific models as a template for the generation of virtual populations. Therefore, ad hoc sensitivity analyses should be conducted whenever developing a new model. The results from this study support the idea that, because of their computationally expensive nature, such sensitivity studies will benefit from the adoption of statistical emulators to produce Sobol points. Conducting this analysis using deterministic OpenSim would have required 30000 model evaluations each for SGM and SSM. With the typical processing times of 30 seconds per simulations this would amount to more than 20 days of continuous computing on single core machines. Conversely, the 50 simulations each for SGM and SSM needed to train the emulator could be run in only 25 minutes, a speedup of 1200%.

Overall, the SA found a correspondence between the muscles deemed as influential for a certain joint and the real anatomical and functional role of those muscles. For what concern SSM, we found a strong dependence of hip JCF on Gluteus Medius $F_{max}$, particularly in the determination of push off peak (also referred to as Peak 2), with a moderate correlation between these parameters. A recently published study showed that the hip loading is significantly affected by modifications in the strength of Gluteus Medius [6, 29]. However, in our study Gluteus Medius' $F_{max}$ had an even larger effect on the resulting knee JCF, particularly on load acceptance peak force (Peak 1), where we also observed a strong negative correlation between the parameters. This can be ascribed to the strict relationship between hip and knee movement and hypothetical crosstalk due to the limited knee motion (1 degree of freedom) that is compensated at the hip level through non-sagittal motion [15]. Knee JCF was also dependent on Tensor Fasciae Latae, Rectus Femoris and Gastrocnemius Medialis $F_{max}$, with the latter having a significant positive correlation with push off peak force. In a previous similar study focussing on the knee joint, Navacchia et al. [18] also found that Gluteus Medius and Gastrocnemius Medialis are major players in the determination of peak contact forces. Interestingly, the correlation coefficients between pennation angle and peak knee JCF reported by them match the R values obtained here for the correlation between peak knee JCF and $F_{max}$ but have opposite sign The pennation angle appears, through its cosine, in the theoretical definition of $F_{max}$ [41]. While this parameter is not explicitly present in our model, it is possible that the

 doi:10.20944/preprints202211.0428.v1



generation of the virtual population for SA drew samples from the parameter space whose effect was equivalent to the sampling of the pennation angle, thus partially explaining the similar correlations observed. However, further refinement of the input space are needed to clarify this aspect. Lastly, Tibialis Posterior and Anterior and Soleus $F_{max}$ had an impact on the determination of ankle JCF, with a strong negative correlation between the $F_{max}$ of soleus and ankle peak force. Interestingly, some unexpected relationships were found too, such as the effect of Gastrocnemius Medialis on the hip or the Rectus Femoris on the ankle. They are likely due to higher order interactions, which cause the effect of one muscle to depend on the value of other muscles' $F_{max}$. This hypothesis should be further investigated in future studies by reducing the number of variables in the sensitivity analysis.

In the case of SGM, hip and knee JCFs were particularly sensitive to the $F_{max}$ of Tensor Fasciae Latae, Sartorius, Rectus Femoris and Gracilis. These muscles do indeed play a role in the actuation of hip and knee joints, particularly in the sagittal and frontal plane. Interestingly, previous sensitivity studies either ignored the role of Gracilis [12, 18] or found its effect negligible towards the estimate of knee JCF [27]. Our results showed a dominant effect of this muscle, particularly on the knee and can be explained by the input $F_{max}$ values associated to Gracilis, which varied by over 35% (second largest variation amongst the analysed muscles) according to literature measurements [30]. This figure was measured over a small cohort of eleven older women and therefore may not be representative of a larger young mixed-gender population. Nonetheless, it offers a meaningful insight on the dynamic behaviour of subject specific MSKMs.

Overall, full personalisation of muscle $F_{max}$ led to minimal improvement of JCF estimates compared to a reduced model (where we only personalised a reduced set of $F_{max}$). In fact, initial peak difference was small, in the order of 0.2 BW, and dropped below 0.1 BW at the hip and knee by just personalising 9 muscles. We observed a localised increase in ankle $\Delta$max and, to a minor extent, in ankle RMSD. They were both caused by a decrease in the performance of the emulated models in late swing. This occurred with the personalisation of Biceps Femoris long head and Tibialis Anterior, which are indeed influential in late swing. Higher order interaction with other muscles are likely to be responsible for this localised small degradation in emulator performance.

This study had some limitations associated to the input dataset. First, the patterns observed in this analysis were specific to the dataset used in this study. In fact, MRI-based $F_{max}$ values for the participant were similar to the values obtained by scaling generic $F_{max}$. A different dataset, with MRI-based $F_{max}$ that deviate more from scaled values, would have probably led to larger discrepancies between a fully-personalised and a reduced models (as suggested by the width of the JCF bands obtained when simulating a virtual population, Figure 2). This hypothesis suggests a component of subject-specificity in the sensitivity of JCFs to $F_{max}$ values. This was also previously observed by Navacchia et al. in [18] when comparing sensitivity outcomes obtained from MSKMs built from three different participants. Despite some inter-subject differences, they also observed significant similarities across participants. A GP-based study including a larger number of participants

 doi:10.20944/preprints202211.0428.v1



could allow the identification of a common set of influential muscles and lead to the generalisation of a model reduction strategy.

Second, we investigated the effect of a single input parameter, and the variation of muscle $F_{max}$ was not complemented with a corresponding anatomical variation in muscle path (i.e. leading to an increased moment arm when increasing cross-sectional area) or other muscle parameters. It is known from the literature that, amongst others, tendon slack length and optimal fibre length can significantly influence the output of a MSKM, but these parameters are not measurable *in-vivo*, and often left to default values even when subject-specific models are developed. This can lead to unphysiological muscle activations and saturations resulting in non-acceptable values of JCF. We did not discard these results but instead included them in the training and validation dataset. Training the emulator to identify unphysiological muscle activations and associated JCF waveforms could be used as an optimisation strategy for model tuning and identification of acceptable values for $F_{max}$, in line with what previously proposed for musculotendon length parameters [7]. Second, MSK simulations of virtual subjects, and consequently the generation of training points for the emulator, relied on a single kinematic trial, whereas different configuration of $F_{max}$ are likely to result in different kinematics and ground reaction force. When additional gait data are not available, adversarial neural network [42] could help in generating artificial walking trials. When kinematics trials are available, either measured or synthetic, they can be incorporated within the model by adding further dimensions to the input parameter space of the emulator. To minimise the known problems that GPs show when scaled to large datasets [21], the Fourier components of the joint kinematics can be used rather than the entire waveforms.

The strategy for muscle personalisation adopted in this study was based on the evaluation of muscle VRSI on the three joints separately, prioritising muscles with high influence over prolonged periods of time. This led to joints having muscles personalised that are not directly relevant to them and caused occasional higher order interactions between muscle $F_{max}$ with the effect of locally increasing the RMSD. Nevertheless, the results stemming from this approach proved in agreement with the literature on MSKMs [6, 7, 18, 29] and the physiology and anatomy of the musculoskeletal system. Alternative personalisation strategies could be explored, for example deriving the order of personalisation as the solution of an optimisation problem which aims at minimising an objective function which accounts for the differences in the outputs between the reduced model and the fully-personalised one. In the model presented here, the discrepancies between reduced and fully-personalised models were extremely low both in terms of RMSD and $\Delta max$, thus corroborating the choice of the personalisation order.

In conclusion, the emulator-based approach for sensitivity analysis and model reduction presented in this study gives results consistent with the existing literature and suggests that $F_{max}$ may be not as relevant as other parameters in determining the dynamic behaviour of personalised MSKM when investigating gait in elderly women. While pursuing a high degree of personalisation of the models leads to more accurate and reliable outputs, cost- and time-related concerns suggest that an appropriate level of

personalisation should be decided according to the question of interest, resolution of data, and other constraints. This study offers, for one subject, a method for estimating which muscles' $F_{max}$ to personalise and according to which order, proving a way of improving model accuracy with relatively low effort.

**Author Contributions:** Conceptualization, I.B. and E.M.; methodology, I.B. and E.M.; software, I.B and E.M..; validation, I.B. and E.M.; formal analysis, I.B. and E.M.; data curation, I.B. and E.M.; writing—original draft preparation, I.B. and E.M.; writing—review and editing, I.B., E.M., A.M. and C.M.; visualization, I.B. and E.M.; supervision, A.M. and C.M.; funding acquisition, A.M. and C.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the EPSRC Frontier Engineering Awards, MultiSim and MultiSim2 projects (Grant Reference Numbers: EP/K03877X/1 and EP/S032940/1) and by the European Commission H2020 programme through the CompBioMed and CompBioMed2 projects (Grant agreements No. 675451 and No. 823712).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The dataset used for the generation of the 200 virtual musculoskeletal models can be found at https://doi.org/10.15131/shef.data.9934055.v3

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

[1] Delp, S. L., Loan, J. P., Hoy, M. G., Zajac, F. E., Topp, E. L., and Rosen, J. M. An interactive graphics-based model of the lower extremity to study orthopaedic surgical procedures **1990**, *IEEE Trans Biomed Eng,* vol. 37, no. 8, pp. 757-767, doi: 10.1109/10.102791.

[2] Pitto, L., Kainz, H., Falisse, A., Wesseling, M., Van Rossom, S., Hoang, H., Papageorgiou, E., Hallemans, A., Desloovere, K., and Molenaers, G. SimCP: A simulation platform to predict gait performance following orthopedic intervention in children with cerebral palsy **2019**, *Front Neurorobot,* vol. 13, p. 54, doi: 10.3389/fnbot.2019.00054.

[3] Montefiori, E., Modenese, L., Di Marco, R., Magni-Manzoni, S., Malattia, C., Petrarca, M., Ronchetti, A., De Horatio, L. T., Van Dijkhuizen, P., Wang, A., Wesarg, S., Viceconti, M., and Mazzà, C. Linking Joint Impairment and Gait Biomechanics in Patients with Juvenile Idiopathic Arthritis **2019**, *Ann Biomed Eng,* vol. 47, no. 11, pp. 2155-2167, doi: 10.1007/s10439-019-02287-0.

[4] Arnold, E. M., Ward, S. R., Lieber, R. L., and Delp, S. L. A Model of the Lower Limb for Analysis of Human Movement **2010**, *Ann Biomed Eng,* vol. 38, no. 2, pp. 269-279, doi: 10.1007/s10439-009-9852-5.

[5] Scheys, L., Loeckx, D., Spaepen, A., Suetens, P., and Jonkers, I. Atlas-based non-rigid image registration to automatically define line-of-action muscle models: a validation study **2009**, *J Biomech,* vol. 42, no. 5, pp. 565-572, doi: 10.1016/j.jbiomech.2008.12.014.

[6] Valente, G., Pitto, L., Testi, D., Seth, A., Delp, S. L., Stagni, R., Viceconti, M., and Taddei, F. Are Subject-Specific Musculoskeletal Models Robust to the Uncertainties in Parameter Identification? **2014**, *PLoS One,* vol. 9, no. 11, p. e112625, doi: 10.1371/journal.pone.0112625.

[7] Modenese, L., Montefiori, E., Wang, A., Wesarg, S., Viceconti, M., and Mazzà, C. Investigation of the dependence of joint contact forces on musculotendon parameters using a codified workflow for image-based modelling **2018**, *J Biomech,* vol. 73, pp. 108-118, doi: 10.1016/j.jbiomech.2018.03.039.

[8] Charles, J. P., Grant, B., D' Août, K., and Bates, K. T. Subject - specific muscle properties from diffusion tensor imaging significantly improve the accuracy of musculoskeletal models **2020**, *J Anat*, vol. 237, no. 5, pp. 941-959, doi: 10.1111/joa.13261.

[9] Arnold, A. S., Salinas, S., Hakawa, D. J., and Delp, S. L. Accuracy of Muscle Moment Arms Estimated from MRI-Based Musculoskeletal Models of the Lower Extremity **2000**, *Comput Aided Surg*, vol. 5, no. 2, pp. 108-119, doi: 10.3109/10929080009148877.

[10] Lenaerts, G., Bartels, W., Gelaude, F., Mulier, M., Spaepen, A., Van der Perre, G., and Jonkers, I. Subject-specific hip geometry and hip joint centre location affects calculated contact forces at the hip during gait **2009**, *J Biomech*, vol. 42, no. 9, pp. 1246-1251, doi: 10.1016/j.jbiomech.2009.03.037.

[11] Benemerito, I., Modenese, L., Montefiori, E., Mazza, C., Viceconti, M., Lacroix, D., and Guo, L. An extended discrete element method for the estimation of contact pressure at the ankle joint during stance phase **2020**, *Proc Inst Mech Eng H J Eng Med*, vol. 234, no. 5, pp. 507-516, doi: 10.1177/0954411920905434.

[12] Carbone, V., Van der Krogt, M., Koopman, H. F., and Verdonschot, N. Sensitivity of subject-specific models to Hill muscle–tendon model parameters in simulations of gait **2016**, *J Biomech*, vol. 49, no. 9, pp. 1953-1960, doi: 10.1016/j.jbiomech.2016.04.008.

[13] Wesseling, M., De Groote, F., Bosmans, L., Bartels, W., Meyer, C., Desloovere, K., and Jonkers, I. Subject-specific geometrical detail rather than cost function formulation affects hip loading calculation **2016**, *Comput Methods Biomech Biomed Engin*, vol. 19, no. 14, pp. 1475-1488, doi: 10.1080/10255842.2016.1154547.

[14] Carbone, V., van der Krogt, M. M., Koopman, H. F. J. M., and Verdonschot, N. Sensitivity of subject-specific models to errors in musculo-skeletal geometry **2012**, *J Biomech*, vol. 45, no. 14, pp. 2476-2480, doi: 10.1016/j.jbiomech.2012.06.026.

[15] Montefiori, E., Hayford, C. F., and Mazzà, C. Variations of lower-limb joint kinematics associated with the use of different ankle joint models **2022**, *J Biomech*, vol. 136, p. 111072, doi: 10.1016/j.jbiomech.2022.111072.

[16] Conconi, M., Montefiori, E., Sancisi, N., and Mazzà, C. Modeling musculoskeletal dynamics during gait: evaluating the best personalization strategy through model anatomical consistency **2021**, *Appl Sci*, vol. 11, no. 18, p. 8348, doi: 10.3390/app11188348.

[17] Martelli, S., Valente, G., Viceconti, M., and Taddei, F. Sensitivity of a subject-specific musculoskeletal model to the uncertainties on the joint axes location **2015**, *Comput Methods Biomech Biomed Engin*, vol. 18, no. 14, pp. 1555-1563, doi: 10.1080/10255842.2014.930134.

[18] Navacchia, A., Myers, C. A., Rullkoetter, P. J., and Shelburne, K. B. Prediction of in vivo knee joint loads using a global probabilistic analysis **2016**, *J Biomech Eng*, vol. 138, no. 3, doi: 10.1115/1.4032379.

[19] Saltelli, A. Making best use of model evaluations to compute sensitivity indices **2002**, *Comput Phys Commun*, vol. 145, no. 2, pp. 280-297, doi: 10.1016/S0010-4655(02)00280-1.

[20] Melis, A., Clayton, R. H., and Marzo, A. Bayesian sensitivity analysis of a 1D vascular model with Gaussian process emulators **2017**, *Int J Numer Method Biomed Eng*, vol. 33, no. 12, p. e2882, doi: 10.1002/cnm.2882.

[21] Williams, C. K. and Rasmussen, C. E., *Gaussian processes in machine learning*. Cambrige, MA: MIT press, 2006.

[22] Chang, E. T. Y., Strong, M., and Clayton, R. H. Bayesian sensitivity analysis of a cardiac cell model using a Gaussian process emulator **2015**, *PLoS One*, vol. 10, no. 6, p. e0130252, doi: 10.1371/journal.pone.0130252.

[23] Han, J., Zhang, X.-P., and Wang, F. Gaussian Process Regression Stochastic Volatility Model for Financial Time Series **2016**, *IEEE J Sel Top Signal Process*, vol. 10, no. 6, pp. 1015-1028, doi: 10.1109/jstsp.2016.2570738.

[24] Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., and Aigrain, S. Gaussian processes for time-series modelling **2013**, *Philos Trans A Math Phys Eng Sci*, vol. 371, no. 1984, p. 20110550, doi: 10.1098/rsta.2011.0550.

[25] Gurchiek, R. D., Ursiny, A. T., and McGinnis, R. S., "Modeling Muscle Synergies as a Gaussian Process: Estimating Unmeasured Muscle Excitations using a Measured Subset," Proceedings of the *Annu Int Conf IEEE Eng Med Biol Soc*, Montreal, QC, Canada, 20-24 July 2020, doi: 10.1109/embc44109.2020.9176232.

[26] Deisenroth, M. P., Fox, D., and Rasmussen, C. E. Gaussian Processes for Data-Efficient Learning in Robotics and Control **2015**, *IEEE Trans Pattern Anal Mach Intell,* vol. 37, no. 2, pp. 408-423, doi: 10.1109/tpami.2013.218.

[27] De Groote, F., Van Campen, A., Jonkers, I., and De Schutter, J. Sensitivity of dynamic simulations of gait and dynamometer experiments to hill muscle model parameters of knee flexors and extensors **2010**, *J Biomech,* vol. 43, no. 10, pp. 1876-1883, doi: 10.1016/j.jbiomech.2010.03.022.

[28] Handsfield, G. G., Meyer, C. H., Hart, J. M., Abel, M. F., and Blemker, S. S. Relationships of 35 lower limb muscles to height and body mass quantified using MRI **2014**, *J Biomech,* vol. 47, no. 3, pp. 631-638, doi: 10.1016/j.jbiomech.2013.12.002.

[29] Altai, Z., Montefiori, E., van Veen, B., A. Paggiosi, M., McCloskey, E. V., Viceconti, M., Mazzà, C., and Li, X. Femoral neck strain prediction during level walking using a combined musculoskeletal and finite element model approach **2021**, *PLoS One,* vol. 16, no. 2, p. e0245121, doi: 10.1371/journal.pone.0245121.

[30] Montefiori, E., Kalkman, B. M., Henson, W. H., Paggiosi, M. A., McCloskey, E. V., and Mazzà, C. MRI-based anatomical characterisation of lower-limb muscles in older women **2020**, *PLoS One,* vol. 15, no. 12, p. e0242973, doi: 10.1371/journal.pone.0242973.

[31] Hicks, J. L., Uchida, T. K., Seth, A., Rajagopal, A., and Delp, S. L. Is my model good enough? Best practices for verification and validation of musculoskeletal models and simulations of movement **2015**, *J Biomech Eng,* vol. 137, no. 2, doi: 10.1115/1.4029304.

[32] Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., Guendelman, E., and Thelen, D. G. OpenSim: open-source software to create and analyze dynamic simulations of movement **2007**, *IEEE Trans Biomed Eng,* vol. 54, no. 11, pp. 1940-1950, doi: 10.1109/TBME.2007.901024.

[33] Steele, K. M., DeMers, M. S., Schwartz, M. H., and Delp, S. L. Compressive tibiofemoral force during crouch gait **2012**, *Gait Posture,* vol. 35, no. 4, pp. 556-560, doi: 10.1016/j.gaitpost.2011.11.023.

[34] "GPy: A Gaussian process framework in python". Available online: http://github.com/SheffieldML/GPy (accessed 14 November 2022).

[35] Coveney, S. and Clayton, R. H. Sensitivity and uncertainty analysis of two human atrial cardiac cell models using Gaussian process emulators **2020**, *Front Physiol,* vol. 11, p. 364, doi: 10.3389/fphys.2020.00364.

[36] Benemerito, I., Narata, A. P., Narracott, A., and Marzo, A. Determining Clinically-Viable Biomarkers for Ischaemic Stroke Through a Mechanistic and Machine Learning Approach **2022**, *Ann Biomed Eng,* vol. 50, no. 6, pp. 740-750, doi: 10.1007/s10439-022-02956-7.

[37] Sobol, I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates **2001**, *Math Comput Simul,* vol. 55, no. 1-3, pp. 271-280, doi: 10.1016/S0378-4754(00)00270-6.

[38] Herman, J. and Usher, W. SALib: An open-source Python library for sensitivity analysis **2017**, *J Open Source Softw,* vol. 2, no. 9, p. 97, doi: 10.21105/joss.00097.

[39] Kainz, H., Wesseling, M., and Jonkers, I. Generic scaled versus subject-specific models for the calculation of musculoskeletal loading in cerebral palsy gait: Effect of personalized musculoskeletal geometry outweighs the effect of personalized neural control **2021**, *Clin Biomech,* vol. 87, p. 105402, doi: 10.1016/j.clinbiomech.2021.105402.

[40] Mohammed, R. O. and Cawley, G. C., "Over-Fitting in Model Selection with Gaussian Process Regression," in *Machine Learning and Data Mining in Pattern Recognition*: Springer, 2017, pp. 192-205.

[41] Fick, R., *Handbuch der Anatomie und Mechanik der Gelenke*. G. Fisher, 1904.

 doi:10.20944/preprints202211.0428.v1



[42] Bicer, M., Phillips, A. T. M., Melis, A., McGregor, A. H., and Modenese, L. Generative deep learning applied to biomechanics: A new augmentation technique for motion capture datasets **2022**, *J Biomech,* vol. 144, p. 111301, doi: 10.1016/j.jbiomech.2022.111301.