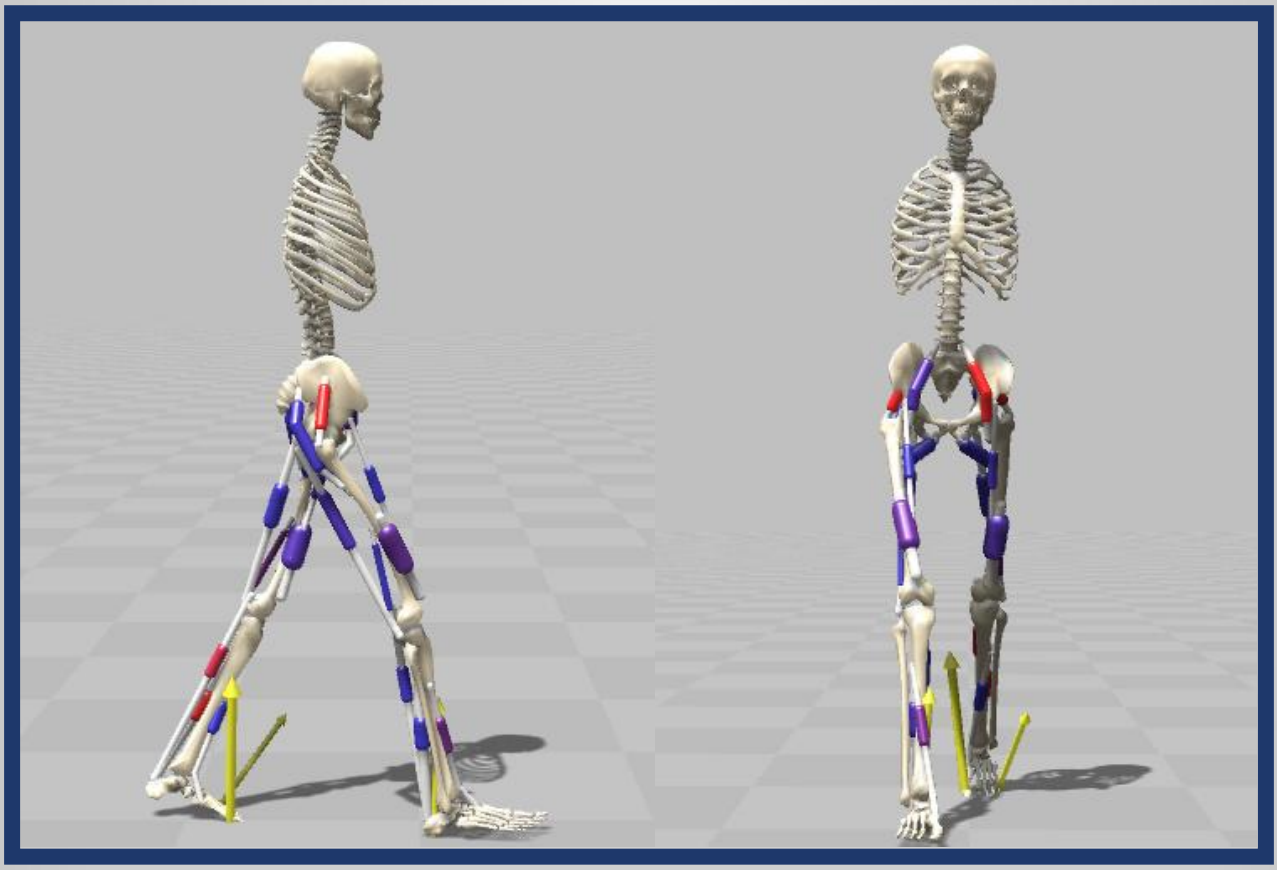
Verification of a 3D Predictive Gait Simulation Framework Using Physiologically Based Objective Functions

Master Thesis

Yuxi Zheng



TU Delft

Verification of a 3D Predictive Gait Simulation Framework Using Physiologically Based Objective Functions

by

Yuxi Zheng
5941482

Supervisors: Eline van der Kruk & Matthew Mulligan

Project duration: July 1, 2025 — January 8, 2026

Faculty: Faculty of Mechanical Engineering, Delft

TU Delft

Abstract—Research on 3D predictive gait simulation remains limited. This study therefore verifies an existing 3D predictive gait simulation framework implemented in SCONE, with the goal that this verified framework can serve as a physiologically plausible modelling and optimisation platform for future related studies. Four optimisation criteria were selected, namely minimising cost of transport, minimising muscle activity, maxmising head instability, and minimising foot-ground impact. These criteria were combined to form a set of objective functions, under which the framework was optimised. The predicted results produced by each objective function were quantitatively compared with experimental data using Pearson r and RMSE/SD, and agreement was evaluated across multiple biomechanical categories, including joint kinematics, ground reaction forces, joint moments and joint powers. The results indicate that, under the optimal objective function, the predictive performance approaches that of the ExpTrack solution that directly tracks experimental data. The framework reproduces sagittal plane hip, knee and ankle angles, vertical and anterior-posterior ground reaction forces, all joint moments and ankle power with strong agreement, but agreement is weaker for hip adduction, medio-lateral ground reaction force, and knee and hip power. Overall, these findings demonstrate the strengths of the framework in reproducing experimental 3D gait, while also revealing limitations in medio-lateral stability control and in the current model and controller settings, providing a basis for targeted improvements in future work.

Keywords—Predictive gait simulation; 3D musculoskeletal model; CMA-ES; Predictive simulation framework verfication

1 Introduction

Predictive simulations of human gait can predict new movement patterns without prescribing experimental joint angles or ground reaction forces (GRFs) in advance [1-5]. Consequently, musculoskeletal modelling and simulation approaches of this type have been widely used to analyse the potential mechanisms underlying movement disorders, to evaluate assistive devices such as orthoses and exoskeletons, and to design rehabilitation training strategies [4, 6-8]. In such predictive simulations, the optimisation algorithm must select an optimal solution from a solution space containing a large number of feasible solutions, so the choice of objective function becomes a central issue, as it determines the optimisation target and, to some extent, reflects assumptions about the optimisation criteria of human gait [9-13].

Previous studies have mainly proposed four physiological optimisation criteria for human gait: minimising energetic cost quantified by cost of transport (CoT), minimising muscle activity (MusAct), maximising head stability (HeadStab), and minimising foot-ground impact (FGImpact). Each of these criteria has a sound theoretical or experimental basis and has been used, either individually or in combination, in different predictive simulation studies. Among them, the most frequently discussed criterion is the CoT [14-19]. A large body of experimental work has shown that humans tend to select combinations of walking speed and step length that minimise energetic cost, so many predictive simulation studies have adopted CoT as the primary or even sole optimisation objective [20-24]. In addition, MusAct encourages the movement task to be performed with lower overall muscle effort, which is commonly interpreted as improving endurance and reducing overall fatigue, and is therefore used as a proxy for anti fatigue demands [5, 22, 25, 26]. HeadStab, implemented as constraints on head linear acceleration, reflects the requirement for vestibular and visual stability [21, 27-29]. and foot-ground impact (FGImpact) is intended to capture control of impact comfort and injury risk [30-34].

In 2D sagittal plane musculoskeletal models, these optimisation criteria have already been examined in a relatively systematic manner. Represented by the work of Veerkamp et al., studies have separately minimised CoT, MusAct, HeadStab, FGImpact and an additional KneeExt term related to the knee extension limit, and then constructed combined objective functions by gradually adding these terms and tuning their weights so that the predicted gait approximates the experimental data as closely as possible [12]. Such studies have both provided a systematic comparison of the roles of different optimisation criteria and clearly demonstrated the strong sensitivity of predictive outcomes to the form of the objective function and the choice of weights. Overall, 2D predictive gait simulations based on these physiological optimisation criteria have developed a relatively complete framework and body of practical experience.

However, these analyses remain confined to 2D sagittal plane models, whereas human gait is inherently 3D. Joint motions in the frontal and transverse planes, such as hip adduction and hip internal-external rotation, together with the associated medio-lateral stability, are equally important for balance and fall risk [35-36]. At present, there is still a lack of systematic predictive gait simulation studies based on these physiological optimisation criteria in full-body 3D musculoskeletal models.

Therefore, this study verifies the plausibility of an existing 3D predictive gait simulation framework implemented in SCONE by quantitatively comparing its predictive results with experimental data. The study aim is to verify, under the reference condition of level walking at self-selected speed in healthy adults, the gait reproduction capability of an existing 3D predictive gait simulation framework implemented in SCONE. The framework comprises the H1922 3D human musculoskeletal model, a reflex based controller, a PD based trunk controller, and CMA-ES optimisation algorithm [37, 38]. Quantitative comparisons with experimental data are used to identify strengths in reproducing experimental 3D gait, to characterise limitations, and to propose directions for improvement. A verified framework is expected to provide a physiologically plausible modelling and optimisation platform for future related studies. To this end, the four physiological optimisation criteria introduced above are used as basic components for constructing the objective function [12]. A range of different objective functions is systematically formulated, and, for each objective function, the agreement between the predictive results and the experimental data is compared across several biomechanical categories, including joint angles, GRFs, joint moments and joint powers. More specifically, the study addresses the following research question:

In the existing SCONE 3D predictive gait simulation framework, which objective function, constructed from combinations of four physiological optimisation criteria, produces predicted gait results with the highest overall agreement with the experimental data? Under this optimal objective function, what are the strengths and limitations of the framework in reproducing 3D gait across different planes and joint DOFs, specifically which measures show strong agreement and which show weaker agreement?

2 Methods

2.1 H1922 3D gait simulation framework

This study is based on an existing 3D predictive gait simulation framework implemented in SCONE [37]. The H1922 3D human musculoskeletal model used in this framework comprised 19 degrees of freedom, including six rigid body degrees of freedom of the pelvis, three rotational degrees of freedom of the lumbar segment, and ten joint degrees of freedom across both legs, namely hip flexion, hip adduction, hip rotation, knee flexion, and ankle flexion. The lower limbs contained 22 Hill type muscle tendon units (11 muscles per leg). For each muscle, the maximum isometric force, optimal fibre length, tendon slack length and pennation angle parameters are listed in Appendix Table A1.

To prevent knee hyperextension, a passive joint limit torque was added at each knee in the H1922 model. This torque was activated when the knee flexion angle was smaller than 5 degrees, with a stiffness of 500 Nm/rad. Foot ground contact was modelled using a Hunt Crossley contact formulation. For each foot, one viscoelastic sphere with a radius of 3 cm was placed under the heel, and two viscoelastic spheres with a radius of 2 cm were placed under the forefoot region to compute the ground reaction forces. In the Hunt Crossley model, the damping coefficient was set to 1 kg/m and the stiffness coefficient was set to 11006.4 N/m. The static and dynamic friction coefficients were set to 0.9 and 0.6, respectively [39].

At the control level, the framework employs a reflex-based neural controller derived from the work of Geyer and Herr to generate muscle excitations for the lower limb muscles [40]. In the original implementation of this framework, the reflex architecture is moderately adapted, and additional reflex pathways to the gluteus medius and adductor magnus are included to improve frontal plane stability control. The detailed reflex structure is listed in Appendix Table A2. To control trunk posture, a PD based trunk controller acted at the lumbar joint level and generated torques in the sagittal, frontal and transverse directions to regulate trunk orientation. All reflex delays were kept identical to those specified in Geyer and Herr's work [40].

The controller contained 93 optimisation variables in total: 16 variables controlling the initial state of the DOFs, 1 variable defining the gait phase, 7 variables specifying the PD and reflex gains for the lumbar joint torques, and the remaining 69 variables specifying reflex gains and offsets for each muscle in different gait phases.

2.2 Optimisation algorithm and simulation settings

All control parameters were optimised using the CMA-ES algorithm [38]. Each optimisation run produced 10 s of predictive gait results. Each optimisation parameter was initialised using the controller's original settings, but the standard deviations were adjusted to obtain stable gait.

To obtain a stable gait solutions, several penalty terms were included in the optimisation: a large penalty was added if a fall

occurred within the 10 s simulation, if the average walking speed was lower than 0.75 m/s, or if ankle flexion exceeded 60 degrees [9]. For each objective function, CMA-ES was run 30 times in parallel with different random seeds, yielding 30 sets of predictive results. The maximum number of generations per optimisation run was set to 2000, and optimisation was terminated earlier if the decrease in the objective function value between two consecutive generations fell below 1×10^{-5}. The trial with the lowest optimisation score was selected as the representative predictive results for that objective function and used for comparison and analysis against the experimental 3D gait data.

2.3 Physiologically-based criteria

Within the H1922 3D gait simulation framework, four physiological optimisation criteria proposed by Veerkamp et al. were adopted as basic components for constructing the objective function [12]. Minimising cost of transport (CoT): based on an Umberger type metabolic energy model [41], the metabolic power of all muscles was integrated over time to obtain the total metabolic energy, which was then divided by the forward displacement of the model centre of mass to represent the metabolic cost per unit distance. Minimising this measure therefore corresponds to minimising energy cost during gait [42]:

$$CoT = \frac{1}{distance * mass} * \int_0^{t_{end}} \left[\sum_{m=1}^{18} \dot{E}_m(t) \right] dt$$

Minimising muscle activity (MusAct): In optimisation, minimising muscle activity is commonly interpreted as encouraging the model to accomplish the movement task with lower overall muscle force, thereby increasing endurance and reducing overall fatigue [5, 22, 25, 26]. For each lower limb muscle, the activation is squared and integrated over time, then the results are summed over all muscles and divided by the forward distance, yielding a total activation² per meter:

$$MusAct = \frac{1}{distance} * \int_0^{t_{end}} [\sum_{m=1}^{22} activation_m(t)^2] dt$$

Mammising head stability (HeadStab): in the 3D model, the acceleration of the head centre of mass is computed in the anterior-posterior (AP), medio-lateral (ML) and vertical directions. The sum of the absolute values in these three directions is used as an instantaneous quantity, which is then integrated over time and divided by the forward distance, reflecting the overall level of head linear acceleration [21, 27-29]:

$$HeadStab = \frac{1}{distance} * \int_0^{t_{end}} [\, |a_x(t)| + |a_y(t)| + |a_z(t)| \,] dt$$

Minimising foot-ground impact (FGImpact): for the 3D GRFs of both feet (vertical, AP and ML), the absolute value of their time derivatives is integrated over time and divided by the forward distance, providing a measure of the rate of change of plantar loading and impact related loading [30-34]:

$$FGImpact = \frac{1}{distance} * \int_0^{t_{end}} [\left| \frac{dGRF_{x,left}(t)}{dt} \right| + \left| \frac{dGRF_{y,left}(t)}{dt} \right| + \left| \frac{dGRF_{z,left}(t)}{dt} \right| + \left| \frac{dGRF_{x,right}(t)}{dt} \right| + \left| \frac{dGRF_{y,right}(t)}{dt} \right| + \left| \frac{dGRF_{z,right}(t)}{dt} \right|] dt$$

2.4 Objective function combinations and weighting

Based on the four criteria described above, different objective functions were constructed to systematically evaluate their influence on predictive gait and to identify an optimal objective function. First, CoT, MusAct, HeadStab and FGImpact were each used as a single objective function in separate optimisations, and the overall agreement between the predictive results and the experimental data was compared for these four single criterion cases. In line with previous studies and the present results [20-24], using CoT alone yielded the best overall agreement, so CoT was treated as a required base criterion in objective functions

constructed from multiple criteria.

All objective functions constructed from multiple criteria took the form "CoT + additional criteria", including CoT+MusAct, CoT+HeadStab, CoT+FGImpact and combinations of CoT with any two or all three of the additional criteria. To avoid differences in physical units causing any single criterion to dominate the optimisation, a constant normalisation factor was predefined for each criterion, and the criterion values were normalised to criterion* so that their magnitudes were comparable around a representative gait solution. The final objective function was written as $J = \sum_i w_i \ criterion_i^*$.

In all objective functions, the weight of CoT was fixed at 1, whereas the weights of the other three criteria were selected from the finite set {0.1, 0.25, 1, 1.5, 5} to construct a series of representative weighting schemes. For each "CoT + additional criteria" objective function form, CMA-ES was run under these discrete weight configurations, and the configuration with the best overall agreement was selected as the final predictive results for that objective function form [12].

2.5 Experimental and predictive data preprocessing

Experimental data were taken from the full body gait dataset published by Van Criekinge et al. [43]. This study used the group mean data and corresponding standard deviations from 29 healthy adults aged 20 to 30 years in the experimental dataset, including 14 males and 15 females, with an age of 25.21 ± 2.85 years, a body mass of 74.24 ± 14.10 kg, and a height of 1.72 ± 0.08 m, collected during level walking at self-selected speed. The variables included lower limb joint angles, three dimensional ground reaction forces, joint moments, and joint powers. The group mean walking speed of these participants was 1.24 ± 0.21 m/s, with a speed range of 1.01 to 1.79 m/s. All data were normalised to the gait cycle and resampled to a standard format spanning 0 to 100 percent of the gait cycle.

For the predictive results, each 10s simulation generated by the optimisation was processed by discarding the first two gait cycles and treating the subsequent five consecutive gait cycles as steady state gait. From these five cycles, the corresponding biomechanical variables were extracted, normalised to 0-100% gait cycle and averaged. The resulting averaged waveforms were taken as the predictive results for that simulation and were used to assess agreement with the experimental group mean data.

2.6 Biomechanical categories, agreement metrics and verification procedure

To evaluate agreement between predictive results and experimental data at multiple levels, variables were grouped into four biomechanical categories: (1) Kinematics: hip flexion, hip adduction, hip internal-external rotation, knee flexion and ankle dorsiflexion; (2) GRFs: vertical, AP and ML GRFs; (3) Joint moments: net hip, knee and ankle joint moments in the relevant planes; (4) Joint powers: hip, knee and ankle powers.

For each objective function, two primary metrics were computed to assess agreement between the predictive results and the experimental data [44-46]: (1) the Pearson correlation coefficient r, calculated between the predictive results and the experimental data, used as the primary agreement metric to reflect similarity in waveform shape and timing pattern; (2) the normalised RMSE (RMSE/SD), defined as the RMSE between predictive results and experimental data divided by the experimental SD, used as an auxiliary metric of relative amplitude error.

The above metrics were first computed at the single variable level, then averaged within each biomechanical category, and finally averaged across all categories to obtain, for each objective function , an overall average r and an overall average RMSE/SD. The predictive average walking speed was recorded at the same time. According to the classification in [47], $r < 0.3$ was interpreted as very weak agreement, 0.3 to 0.5 as weak, 0.5 to 0.7 as moderate and $r > 0.7$ as strong. In the present study, Pearson r served as the primary metric for evaluating agreement between predictive results and experimental data, whereas RMSE/SD was used as an auxiliary metric to compare amplitude errors across different objective functions.

On this basis, verification was carried out in two steps. (1) Single criterion analysis: the overall average r (primary) and overall average RMSE/SD (auxiliary) of the four single objective functions CoT, MusAct, HeadStab and FGImpact were compared to evaluate the effect of minimising each criterion individually and to confirm the rationale for treating CoT as the baseline criterion in multi objective optimisation. (2) CoT-based combination analysis: for all "CoT + additional criteria" objective function combinations and their discrete weighting schemes, CMA-ES was run, overall average r and RMSE/SD were computed, and the combination with the highest r and lowest RMSE/SD was selected as the optimal objective function. Under the optimal objective

function, subsequent analyses primarily used Pearson r within each biomechanical category to characterise the framework's reproduction capabilities and limitations across different planes and degrees of freedom.

To assess whether, under the optimal objective function, the framework already approaches the upper performance bound of predictive simulation, an additional tracking objective function (ExpTrack) was constructed. Under ExpTrack, CMA-ES directly tracked the experimental data by minimising RMSE/SD between predictive results and experimental data. The resulting predictive results were regarded as the upper performance bound achievable with this framework and were used for comparison with the results obtained under the optimal objective function combination [12].

In addition, agreement in muscle activation between the predictive results under the optimal objective function and the experimental data was also assessed in this study, but only as a reference. Since the dataset in [43] lacks corresponding EMG measurements, the experimental EMG data were taken from [48]. Specifically, group mean EMG profiles were obtained from 20 healthy adults aged 22 to 72 years, including 9 males and 11 females, with an age of 43.1 ± 15.4 years, a body mass of 68.5 ± 15.8 kg, and a height of 1.71 ± 0.10 m, collected during level walking at a comfortable speed.

3 Results

Video recordings of the predicted gait simulations for all objective functions are available at: https://youtu.be/7e_MZzK5t0A

3.1 Single-criterion simulations

This study first performed optimisation using CoT, MusAct, HeadStab and FGImpact as single objective functions. The results showed that all four single criteria were able to generate 3D predictive gait, as illustrated in Figure 1. However, the overall average Pearson r ranged only from 0.59 to 0.66, which is below 0.7, indicating limited agreement between the predicted gait and the experimental data when a single criterion was used. Consistent with previous studies [20-24] and the present results, the objective function based on CoT alone achieved the best overall performance in terms of overall average r and overall average RMSE/SD, with an overall average r of 0.67 and an RMSE/SD of 1.75. Therefore, CoT was treated as a required base criterion in the subsequent objective functions constructed from multiple criteria.

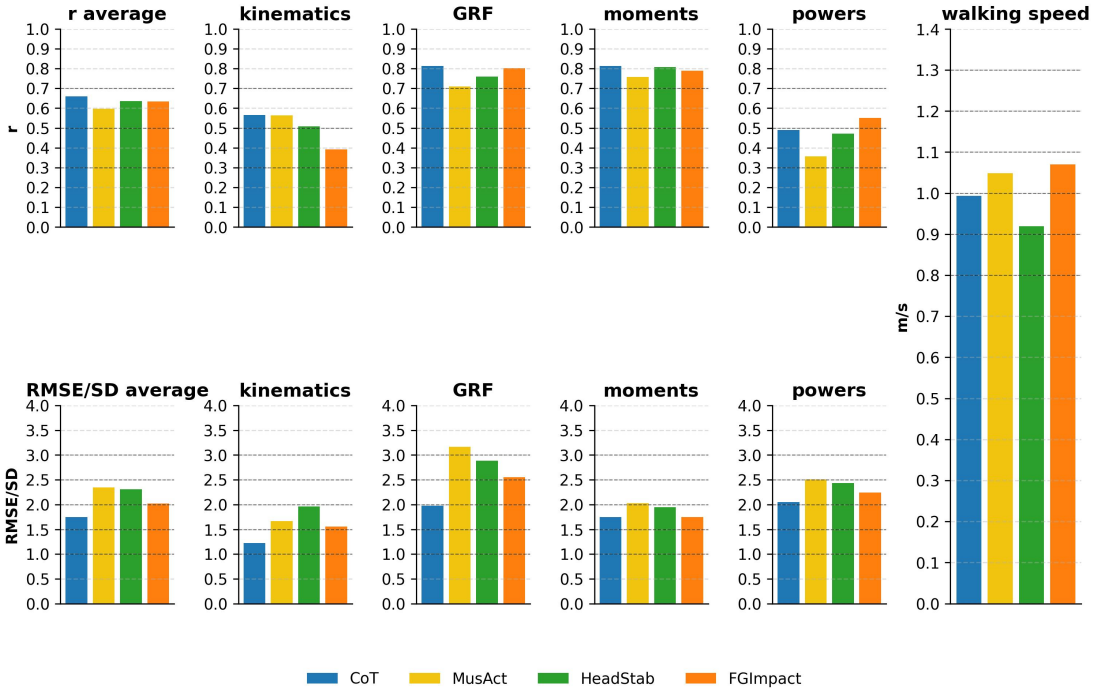


Figure 1. Pearson r and RMSE/SD between results predicted by minimising four criteria and the experimental data.

3.2 CoT-based objective function combinations and optimal combination

After confirming CoT as the baseline criterion, a systematic exploration was carried out of objective function of the form "CoT + additional criteria". When evaluating agreement, overall average Pearson r was used as the primary metric, and overall

average RMSE/SD was used as a secondary metric. Following the grading scheme in [47], r values greater than 0.7 are typically considered to indicate a strong correlation. Accordingly, this study treated an overall average r greater than 0.7 as the main criterion for strong agreement between predictive results and the experimental data. RMSE/SD was mainly used to compare the relative magnitude of amplitude errors across different objective functions, rather than to apply a strict absolute threshold. CoT was treated as a required criterion, and MusAct, HeadStab and FGImpact were added to form different objective functions. For each added criterion, the weight was selected from the discrete set {0.1, 0.25, 1, 1.5, 5}. In total, 215 CoT based optimisation runs were completed. For each objective function constructed from multiple criteria, the representative weight settings were retained based on the highest overall average r together with relatively low average RMSE/SD. The results for these objective functions are shown in Figure 2.

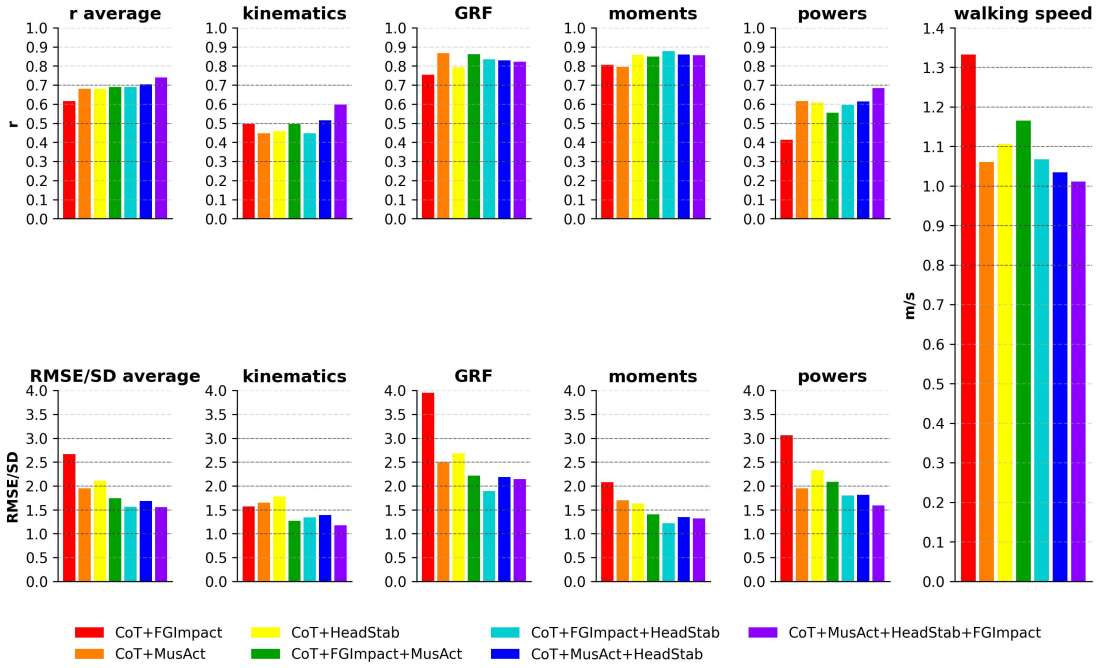


Figure 2. Pearson r and RMSE/SD between results predicted by minimising different objective function combinations and the experimental data.

Among all objective funcions, the best overall agreement was obtained with the objective funcion containing CoT, MusAct, HeadStab and FGImpact: CoT+MusAct+HeadStab+FGImpact. The details of this objective function combination are listed in Table 1, and the waveform profiles of the biomechanical variables under the optimal objective function are shown in Figure 3. Across the biomechanical categories, the average r values for kinematics, GRFs, joint moments and joint powers were 0.60, 0.82, 0.86 and 0.68, respectively, while the corresponding average RMSE/SD values were 1.17, 2.14, 1.32, and 1.59, respectively. The overall average r was 0.74 (the highest among all objective functions) and the overall average RMSE/SD was 1.56 (the lowest among all objective functions). According to the agreement classification adopted in this study, this overall average r falls within the strong range ($r > 0.7$). The corresponding walking speed was 1.01 m/s, which falls within the experimental range of 1.01 to 1.79 m/s, but is lower than the experimental group mean walking speed of 1.24 m/s.

Table 1. Overview of optimal objective function. Normalization factor scales each criterion to a comparable magnitude (criterion* = criterion / factor). Weighting is the dimensionless coefficient w_i applied to each normalised term in the combined objective.

Criteria	Normalization factor	Criterion's optimal objective function weighting
CoT	5.0225	1
FGImpact	6.5831	1
MusAct	1.005	1.5
HeadStab	3.3772	1.5

To further assess the extent to which this optimal objective funcion had already exploited the predictive capability of the framework, an ExpTrack simulation was performed. Under ExpTrack, the overall average r was 0.78 and the overall average RMSE/SD was 1.36, whereas for the optimal objective funcion they were 0.74 and 1.56, respectively (Figure 4). In terms of Pearson r, the optimal objective function achieved performance that was very close to ExpTrack. Since ExpTrack directly tracks the experimental data, its results can be regarded as an upper bound of the agreement achievable under the current framework consisting of the H1922 model, the controllers, and CMA-ES optimisation algorithm. Therefore, CoT+MusAct+HeadStab+FGImpact objective funcion can be considered the optimal objective function identified in this framework using this pesent approach.

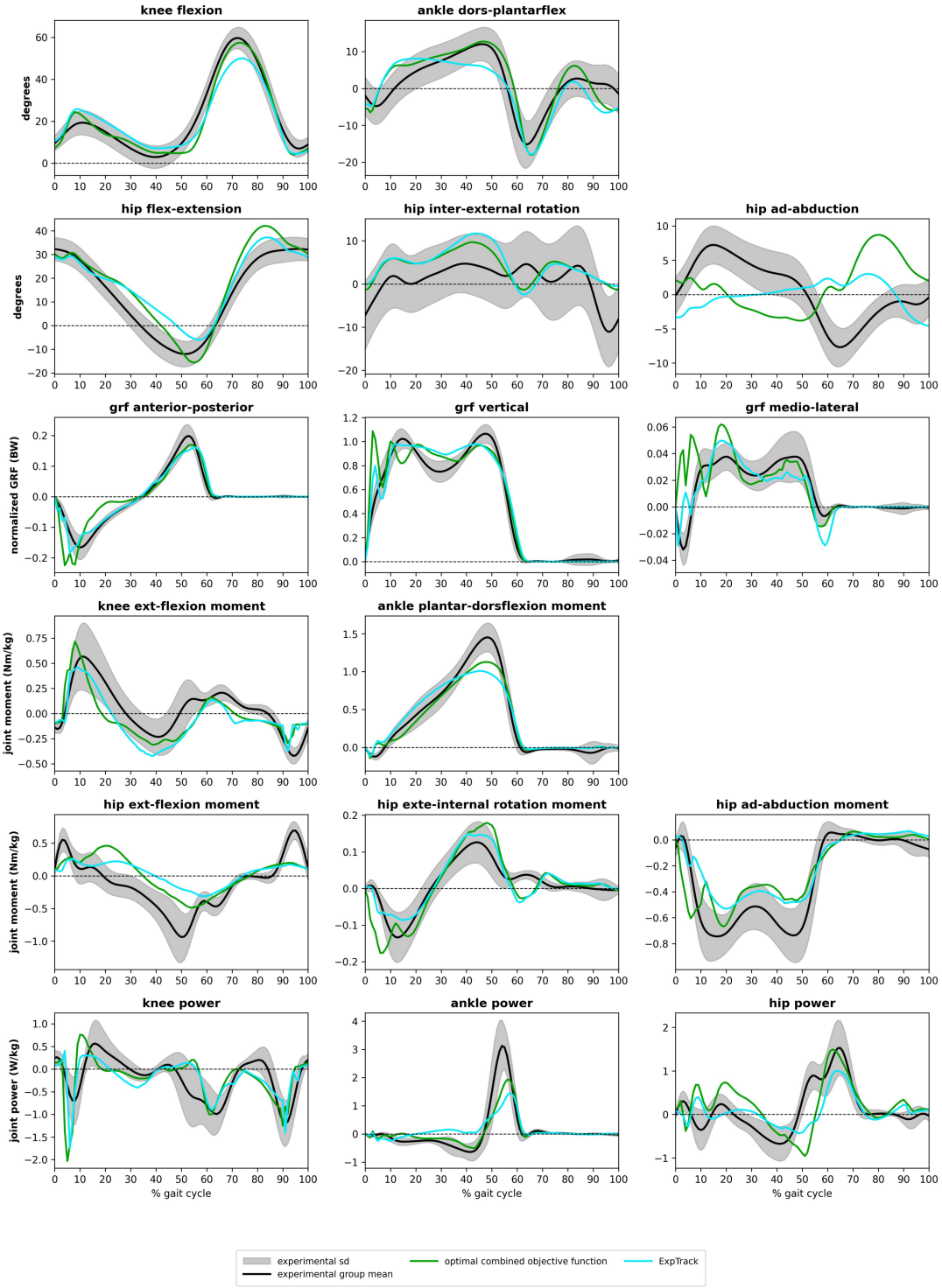


Figure 3. Comparison of waveforms predicted by the optimal objective function combination and by ExpTrack with the experimental data.

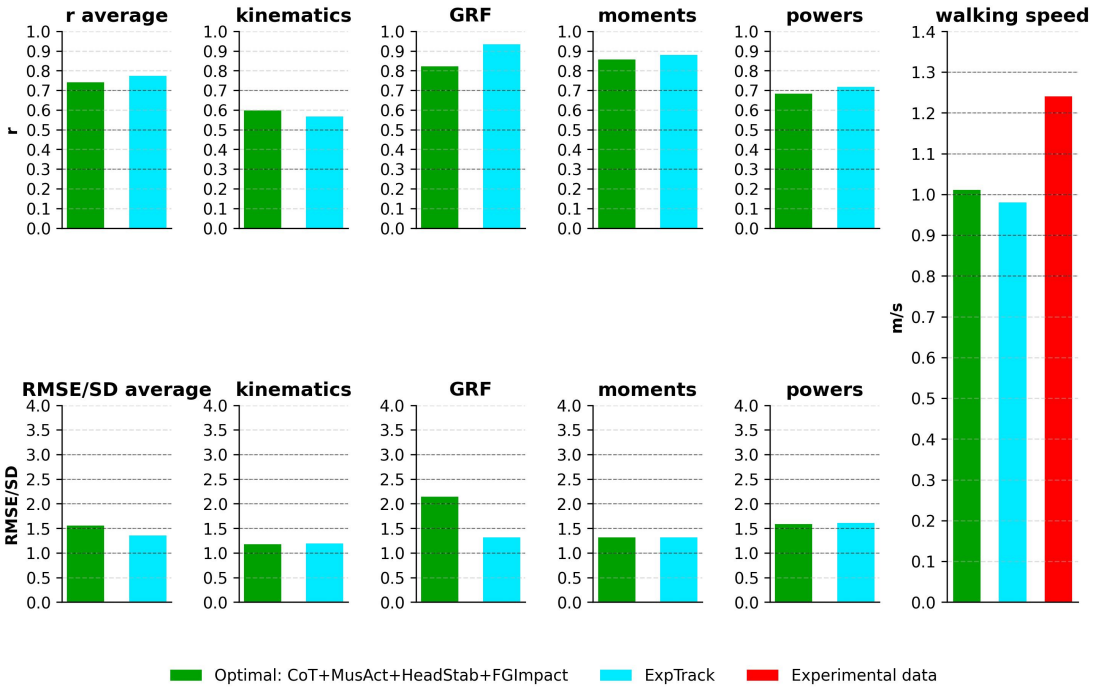


Figure 4. Pearson r and RMSE/SD between results predicted by the optimal objective function combination and by ExpTrack and the experimental data.

3.3 The agreement between the optimal predictive results and the experimental data

After identifying the optimal objective function combination, the framework's reproduction strengths and limitations across different planes and joint degrees of freedom were further analysed based on Pearson r within each biomechanical variables. The results are shown in Figure 5.

Under the optimal objective function, in the sagittal plane, the Pearson r values for hip flexion, knee flexion and ankle flexion angles were 0.91, 0.98 and 0.96, respectively. The r values for vertical and anterior-posterior (AP) GRFs were 0.91 and 0.96, respectively. The r values for the primary sagittal plane joint moments, namely hip flexion, knee flexion and ankle flexion moments, were 0.74, 0.77 and 0.99, respectively. As all these r values exceed 0.7, they indicate strong agreement, suggesting that under the reference condition the framework reproduces the waveform shape and timing of sagittal plane joint kinematics, vertical and AP GRFs, and the corresponding joint moments with high fidelity, capturing key features of healthy gait.

In the frontal and transverse planes, the hip adduction angle showed a negative correlation with the experimental data, with an r of −0.43, whereas the hip rotation angle showed moderate agreement, with an r of 0.58. The medio-lateral (ML) ground reaction force also showed moderate agreement, with an r of 0.59. In contrast, the agreement for joint moments was substantially higher: the hip adduction moment and hip rotation moment achieved strong agreement, with r values of 0.88 and 0.89, respectively.

For joint power, ankle power showed strong agreement, with an r of 0.93, indicating that the predicted patterns of energy generation and absorption at the ankle were very close to the experimental data. In comparison, knee power and hip power showed only moderate agreement, with r values of 0.54 and 0.58, respectively.

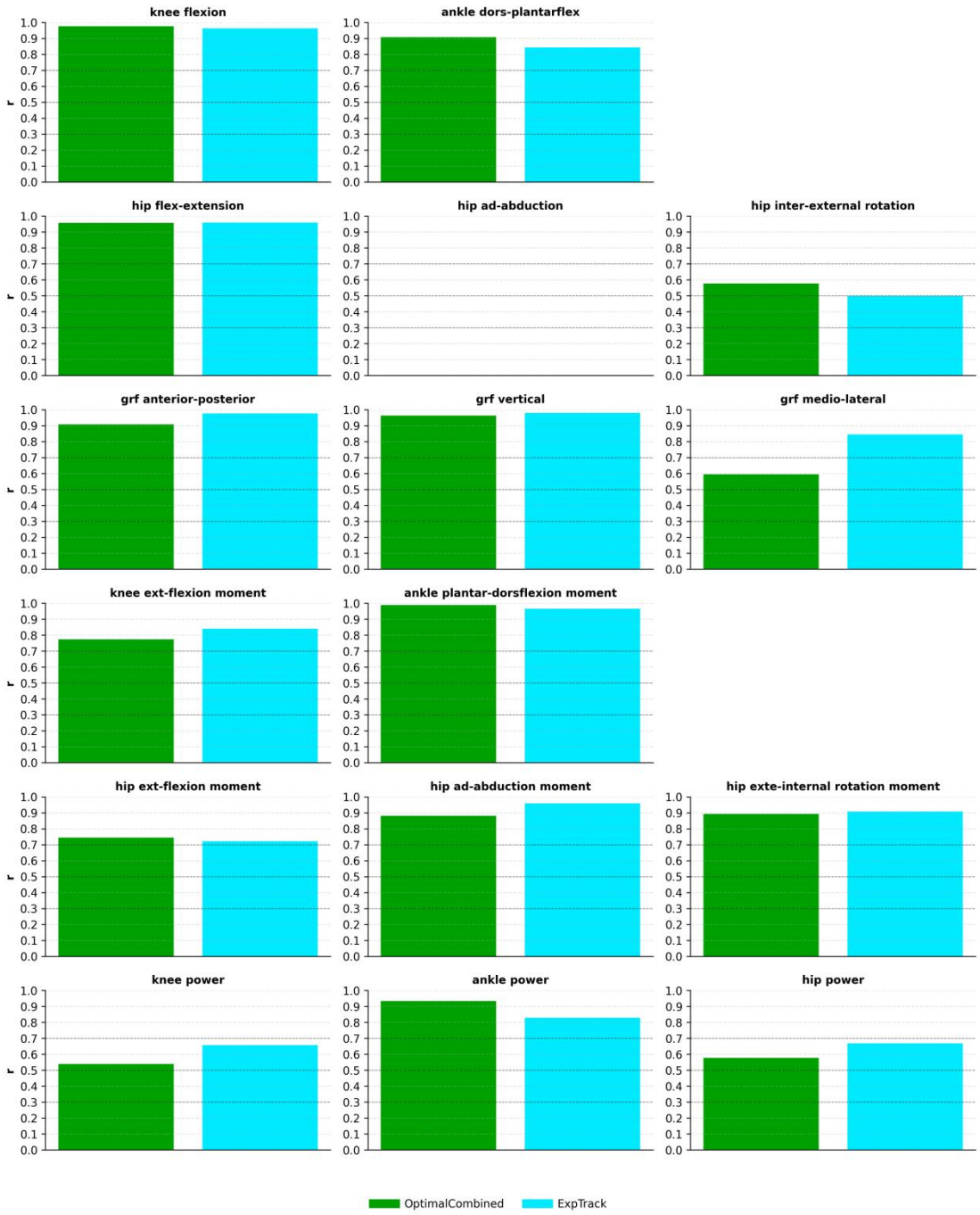


Figure 5. Pearson r between predictive results obtained with the optimal combination and ExpTrack and the experimental data. For hip ad-abduction, the absence of bars indicates negative r values and therefore very poor agreement.

3.4 Muscle activation

Muscle activation was not included as a criterion when selecting the optimal objective function. The main reason is that EMG amplitude is highly dependent on the normalisation procedure and measurement conditions, and therefore the experimental EMG and the framework predicted activations are not directly comparable in terms of absolute amplitude. Nevertheless, Pearson r between the predicted muscle activations and the experimental EMG was computed as a reference, as shown in Figure 6 and 7. In Figure 6, to enable a fairer waveform comparison, both the experimental EMG and the predicted muscle activations were peak normalised within the gait cycle for each muscle. As a result, this figure focuses on the timing and shape of the activation profiles, while reducing the influence of differences in absolute amplitude scaling.

Figure 6 shows that, for both the optimal objective function and ExpTrack, the r values for muscles such as rectus femoris, gluteus maximus, hamstrings, vasti, gastrocnemius and soleus were mostly below 0.7 and were even negative in some cases. The

overall average r values were 0.32 and 0.28, respectively, far below 0.7. These results indicate that the predicted muscle activation profiles show at most a broadly similar trend to the experimental EMG during certain phases, and that the overall agreement is very weak.

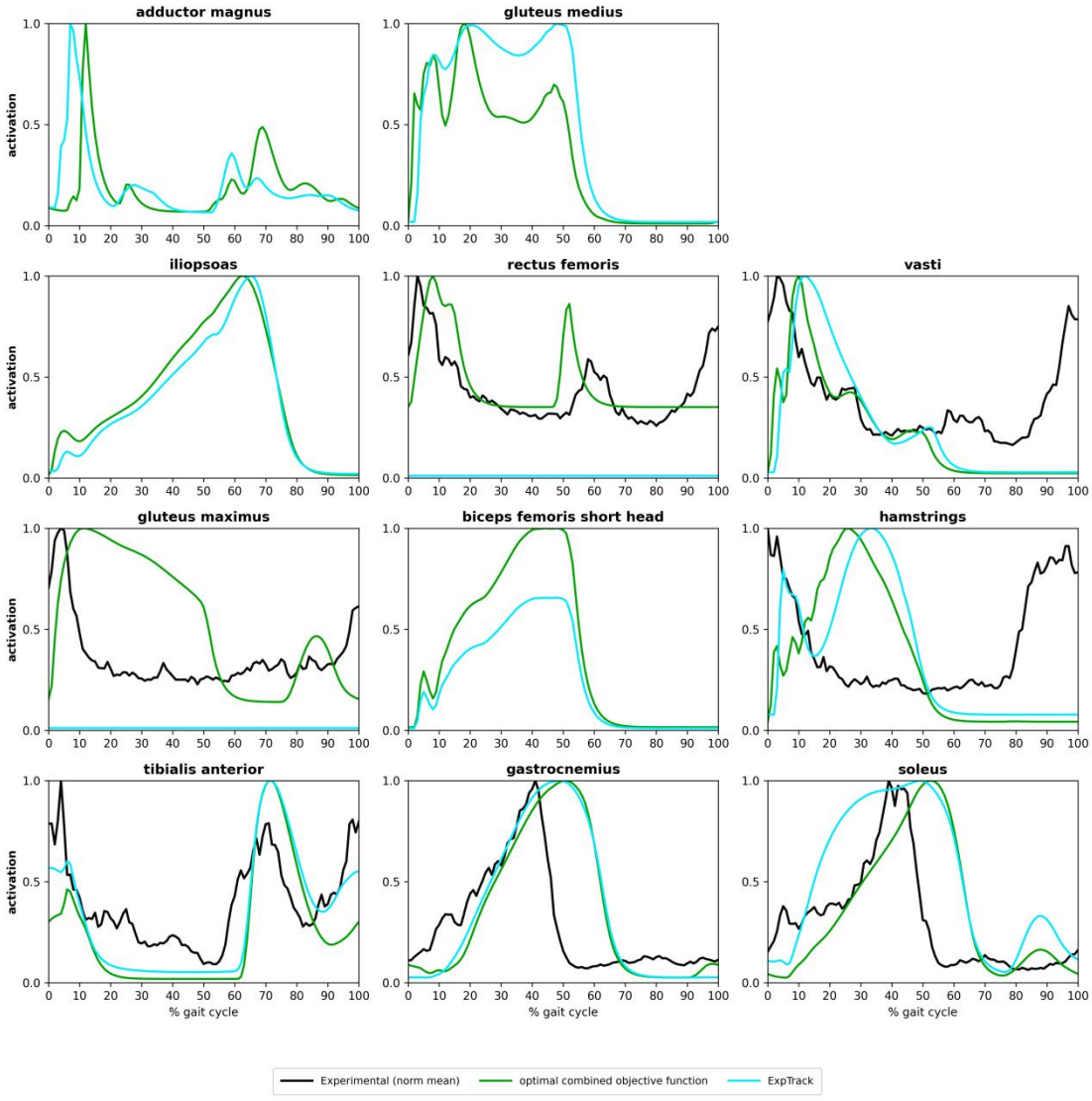


Figure 6. Comparison of muscle activation predicted by the optimal objective function combination and by ExpTrack with the experimental data.

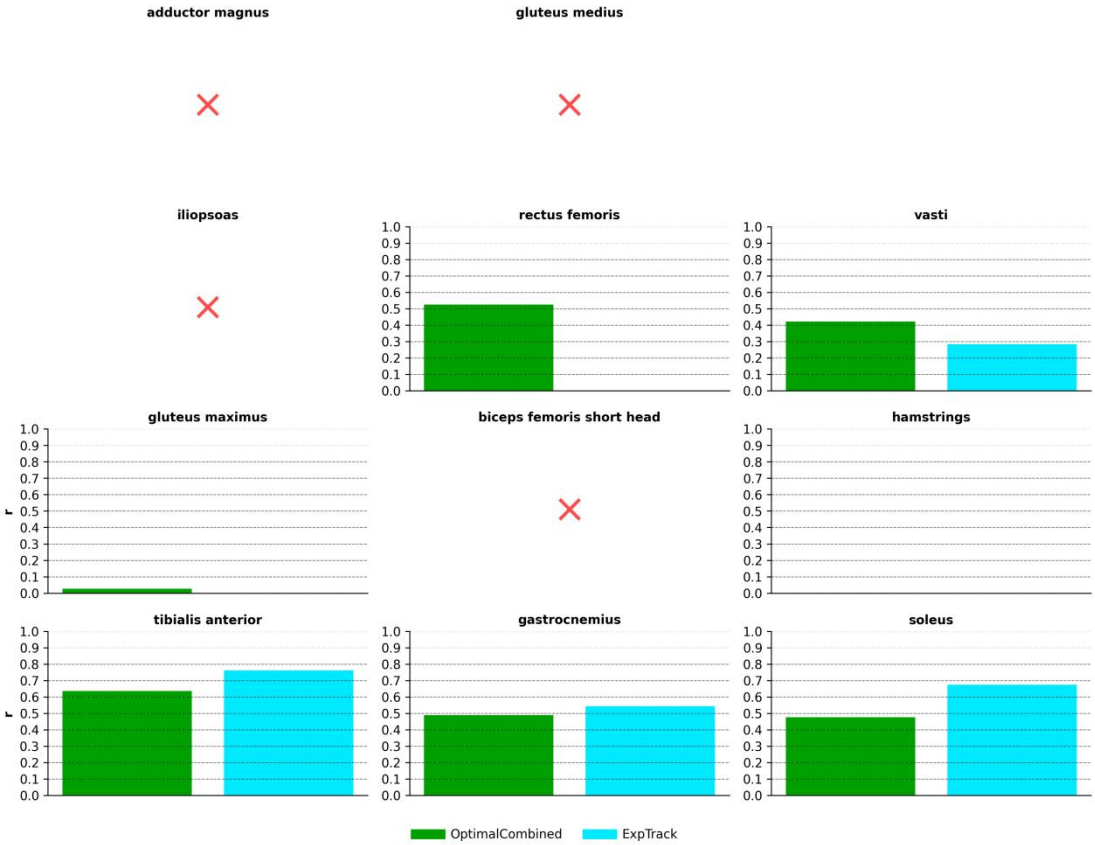


Figure 7. Pearson r between muscle activation predicted by the optimal objective function combination and by ExpTrack and the experimental data. Crosses indicate missing experimental data for which r could not be computed. For rectus femoris, gluteus maximus and the hamstrings, the absence of bars indicates r values below zero and therefore extremely poor agreement.

4 Discussion

4.1 Optimal objective function

Based on the results of this study, under the reference condition of level walking at self-selected speed in healthy adults and without explicitly prescribing experimental tracking targets, the objective funcion contructed from CoT+MusAct+HeadStab+FGImpact can be regarded as a physiologically plausible optimal objective function for the framework consisting of the H1922 model, the reflex-based controllers, and CMA-ES optimisation algorithm. This objective function yields strong agreement between the predictive results and the experimental data and is capable of generating 3D predictive gait simulations that reproduce experimental gait with high fidelity.

4.2 Reproduction capabilities and limitations across planes and DOFs

From the results, the Pearson r between the predicted hip adduction angle and the experimental data is negative, indicating extremely poor agreement. Figure 3 shows that, for both the optimal objective function and the direct tracking optimisation (ExpTrack), the hip adduction angle waveform is almost opposite in shape to the experimental group mean. This phenomenon is likely related to the way medio-lateral (ML) stability is represented in the current H1922 framework. In this model, ML stability is primarily provided by two muscles: gluteus medius and adductor magnus. During the stance phase (0 to 50% gait cycle), the stance leg gluteus medius remains highly activated for a prolonged period (Figure 6) to elevate the contralateral side of the pelvis and create clearance for swing of the opposite leg. In the model, this tends to manifest as pronounced hip abduction on the stance side, rather than the hip adduction observed in the experimental data. As a result, the hip adduction angle waveform becomes reversed in shape relative to the experimental group mean, leading to a negative r.

For hip rotation angle, the r value of 0.58 falls within the moderate agreement range. However, in actual gait, hip rotation exhibits substantial inter subject variability [43]. As shown in Figure 3, the predictive hip rotation curves remain within the

experimental group mean ± 1 SD range for most of the gait cycle, and the timing of the main peaks is reasonably close to that of the experimental data. Therefore, although the Pearson r does not reach the strong range, the reproduction of hip rotation angle by the current framework can still be considered acceptable when waveform shape and acceptable variability are taken into account.

For ML GRF, the r value of 0.59 also corresponds to moderate agreement. On the one hand, the absolute magnitude of ML GRF is much smaller than that of vertical and AP GRFs, making it more susceptible to measurement noise and trial to trial or inter subject variability in the experimental data. On the other hand, ML stability in the current H1922 model relies mainly on a limited set of hip muscles (gluteus medius and adductor magnus) and a relatively simple trunk PD controller, without including additional muscles that contribute to lateral control of the pelvis and trunk. This simplification allows the framework to generate a broadly reasonable trend in ML GRF, but makes it difficult to match the finer amplitude modulation observed experimentally, leading to only moderate agreement in the ML direction.

For joint power, Figure 3 shows that the overall trends of the predicted knee and hip power are broadly consistent with the experimental data, but differences remain in amplitude and fine scale waveform shape. Since joint power equals joint moment multiplied by joint angular velocity, the weaker agreement in hip power is mainly due to the weaker agreement in hip adduction and hip rotation in the frontal and transverse planes. Weaker kinematics agreement leads to weaker agreement in joint angular velocity. It is also noteworthy that, although both knee flexion angle and knee flexion moment show strong agreement, the agreement for knee power is still clearly lower. This is because power is not an independently measured quantity but is computed as the product of joint moment and angular velocity, $P(t) = M(t)\omega(t)$, where $\omega(t) = d\theta(t)/dt$. In the predictive simulations of this study, the time step is very small, with $dt = 0.01s$. As a result, any error between the predicted and experimental joint angle increment $d\theta(t)$ can be amplified by the division by dt, effectively scaling the error by a factor of 100. Therefore, angular velocity, as the first derivative of joint angle, is highly sensitive to small phase shifts and local waveform differences. In addition, errors in the predicted moment and angular velocity can be further amplified through the product operation. For example, writing the predicted joint moment and angular velocity as $M_s = M_e + \Delta M$ and $\omega_s = \omega_e + \Delta\omega$, the predicted joint power becomes $P_s = M_e\omega_e + M_e\Delta\omega + \omega_e\Delta M + \Delta M\Delta\omega$. This expansion shows that the power error terms $\Delta\omega$ and ΔM are amplified by the multiplicative factors. Consequently, joint power is often a more stringent metric than joint angle or moment. Even when angle and moment exhibit high agreement, the agreement of power can still decrease substantially.

4.3 Muscle activation agreement

This study also compared the similarity in waveform shape between the predictive results and the experimental data at the muscle activation level by computing Pearson r. Since the public dataset reported by Van Criekinge et al. does not include EMG signals corresponding to the muscles in the H1922 model [43], experimental EMG profiles for rectus femoris, vasti, gluteus maximus, hamstrings, tibialis anterior, gastrocnemius, and soleus were taken from the literature [48] as a reference. The two sources are not fully matched in terms of subject cohort. In addition, the experimental EMG dataset does not cover all muscles represented in the model. Moreover, EMG amplitude depends strongly on the normalisation method and measurement conditions and is therefore not directly comparable to the framework predicted activations in absolute magnitude [49-50], so RMSE/SD was not computed for this category. To focus the comparison on waveform shape rather than amplitude scaling, both the predicted activations and the experimental EMG were peak normalised within the gait cycle for each muscle before Pearson r was calculated. Based on these considerations, muscle activation was not included in the criteria for selecting the optimal objective function, and Pearson r was used only as a qualitative reference to indicate potential limitations of the framework in reproducing muscle activation patterns.

The results show that Pearson r was low for most muscles, and was even negative for some, with an overall average r far below 0.7. This indicates that the predicted muscle activations exhibit only broadly similar trends to the experimental EMG for a small number of muscles, as shown in Figure 6, and that the overall agreement is weak. Under the current data conditions, the framework cannot satisfactorily reproduce the experimental EMG. However, this conclusion remains uncertain because the experimental EMG is not fully matched to the model muscle set, and because the EMG acquisition and processing pipeline, including electrode placement, filtering, rectification, envelope extraction, and normalisation, cannot be made fully consistent with the settings used in this study. Therefore, the main value of this analysis is to suggest potential limitations of the framework at the

muscle level, rather than to provide strong verification of performance in this category. Future work aiming for a more rigorous verification of muscle activation should use an EMG dataset that closely matches the model muscle set and adopts a normalisation scheme that is more consistent with the framework settings.

4.4 Limitations and future work

Several limitations of this study should be noted. First, verification of the framework was carried out only under the reference condition of level walking at self-selected speed in healthy adults. The generalisability of this framework to other walking speeds, slopes or pathological gait patterns has not yet been examined and will require further validation. Second, to reduce optimisation time, the weights of the objective function were searched only within the finite discrete set {0.1, 1, 1.5}, rather than being systematically explored in a continuous weight space using bilevel optimisation or Monte Carlo based approaches. The optimal objective funcion reported in this study should therefore be interpreted as the best solution found within the current method, rather than being guaranteed as a global optimum. Third, under the optimal objective function, the predicted walking speed was 1.01 m/s. Although this value falls within the experimental speed range of 1.01 to 1.79 m/s, it lies at the lower bound of the range and is below the experimental group mean of 1.24 m/s. Because the predicted walking speed corresponds only to the lower end of the experimental walking speed distribution, comparing the predictive results with the experimental group mean does not constitute a strictly speed matched comparison. Since joint angles, ground reaction forces, joint moments and joint powers are all highly sensitive to walking speed, this lower speed may act as a potential confounding factor. Therefore, for biomechanical variables with relatively low agreement, including medio-lateral ground reaction force, hip adduction, hip rotation, hip power and knee power, the discrepancies between predictions and experimental data may partly arise from the speed mismatch rather than from limitations of the framework itself. Future work could address this issue by imposing a strict target speed constraint or adding a speed tracking term, or by selecting a speed matched experimental data, with a group mean walking speed close to the predicted value, when recomputing agreement metrics, in order to more rigorously isolate the effect of walking speed.

Future work should therefore further validate and extend the present conclusions by performing more systematic weight searches and testing the framework across multiple tasks and conditions, ideally using validation scenarios that match walking speed.

Conclusion

By comparing four physiologically motivated optimisation criteria, this study identified CoT+MusAct+HeadStab+FGImpact as the optimal objective function. Under this objective function, the predictive results achieved an overall average r of 0.74 and an overall RMSE/SD of 1.56. These values are close to those obtained with ExpTrack, a tracking based objective function, with r of 0.77 and RMSE/SD of 1.36, suggesting that the predictive performance of this objective function approaches the upper performance bound achievable within the current framework. Under the optimal objective function, the framework showed strong agreement with the experimental data for sagittal plane kinematics, vertical and AP GRFs, all joint moments, and ankle power. This indicates the strength of the framework in reproducing 3D gait. In contrast, only moderate agreement was observed for hip adduction angle, ML GRF, knee power and hip power, suggesting that the current model still involves simplifications in ML stability control and in certain aspects of energy transfer. These measures highlight the limitations of the framework and point to potential improvements, including expanding the muscle set, for example by adding more trunk and hip surrounding muscles, and refining the controller. At the muscle level, the experimental EMG data do not fully match the model muscle set, and agreement between predicted activations and experimental EMG was very low for most muscles. This suggests potential limitations in reproducing muscle activation patterns, although more rigorous verification will require future studies using EMG datasets that better match the model muscle set and are processed under more consistent conditions.

Reference

[1] Anderson, F. C., & Pandy, M. G. (2001). Dynamic optimization of human walking. J. Biomech. Eng., 123(5), 381-390.

[2] Ackermann, M., & Van den Bogert, A. J. (2010). Optimality principles for model-based prediction of human gait. Journal of biomechanics, 43(6), 1055-1060.

[3] Song, S., & Geyer, H. (2013, July). Generalization of a muscle-reflex control model to 3D walking. In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 7463-7466). IEEE.

[4] Ong, C. F., Geijtenbeek, T., Hicks, J. L., & Delp, S. L. (2019). Predicting gait adaptations due to ankle plantarflexor muscle weakness and contracture using physics-based musculoskeletal simulations. PLoS computational biology, 15(10), e1006993.

[5] Falisse, A., Serrancolí, G., Dembia, C. L., Gillis, J., Jonkers, I., & De Groote, F. (2019). Rapid predictive simulations with complex musculoskeletal models suggest that diverse healthy and pathological human gaits can emerge from similar control strategies. Journal of The Royal Society Interface, 16(157), 20190402.

[6] Wang, Y., Liu, Z., Zhu, L., Li, X., & Wang, H. (2020, December). An impedance control method of lower limb exoskeleton rehabilitation robot based on predicted forward dynamics. In 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) (pp. 1515-1518). IEEE.

[7] Ratnakumar, N., & Zhou, X. (2021, August). Optimized torque assistance during walking with an idealized hip exoskeleton. In International design engineering technical conferences and computers and information in engineering conference (Vol. 85376, p. V002T02A009). American Society of Mechanical Engineers.

[8] Waterval, N. F. J., Veerkamp, K., Geijtenbeek, T., Harlaar, J., Nollet, F., Brehm, M. A., & van der Krogt, M. M. (2021). Validation of forward simulations to predict the effects of bilateral plantarflexor weakness on gait. Gait & Posture, 87, 33-42.

[9] van den Bogert AJ, Blana D, Heinrich D. Implicit methods for efficient musculoskeletal simulation and optimal control. Proc Inst Mech Eng H. 2011;225(10):1214–1226.

[10] Dembia CL, Bianco NA, Falisse A, Hicks JL, Delp SL. OpenSim Moco: musculoskeletal optimal control. PLoS Comput Biol. 2020;16(12):e1008493.

[11] Kuo AD, Donelan JM, Ruina A. Energetic consequences of walking like an inverted pendulum: step-to-step transitions. Exerc Sport Sci Rev. 2005;33(2):88–97.

[12] Veerkamp K, Ong CF, Geijtenbeek T, van der Krogt MM. Evaluating cost function criteria in predicting healthy gait. J Biomech. 2021;123:110524.

[13] Miller RH. A comparison of muscle energy models for simulating human walking in three dimensions. J Biomech. 2014;47(6):1373–1381.

[14] Abram, S. J., Selinger, J. C., & Donelan, J. M. (2019). Energy optimization is a major objective in the real-time control of step width in human walking. Journal of biomechanics, 91, 85-91.

[15] Bertram, J. E., & Ruina, A. (2001). Multiple walking speed–frequency relations are predicted by constrained optimization. Journal of theoretical Biology, 209(4), 445-453.

[16] Maxwell Donelan, J., Kram, R., & Arthur D, K. (2001). Mechanical and metabolic determinants of the preferred step width in human walking. Proceedings of the Royal Society of London. Series B: Biological Sciences, 268(1480), 1985-1992.

[17] Minetti, A. E., Moia, C., Roi, G. S., Susta, D., & Ferretti, G. (2002). Energy cost of walking and running at extreme uphill and downhill slopes. Journal of applied physiology.

[18] Selinger, J. C., O’Connor, S. M., Wong, J. D., & Donelan, J. M. (2015). Humans can continuously optimize energetic cost during walking. Current Biology, 25(18), 2452-2456.

[19] Zarrugh, M. Y., Todd, F. N., & Ralston, H. J. (1974). Optimization of energy expenditure during level walking. European journal of applied physiology and occupational physiology, 33(4), 293-306.

[20] Anderson, F. C., & Pandy, M. G. (2001). Dynamic optimization of human walking. J. Biomech. Eng., 123(5), 381-390.

[21] Dorn, T. W., Wang, J. M., Hicks, J. L., & Delp, S. L. (2015). Predictive simulation generates human adaptations during loaded and inclined walking. PloS one, 10(4), e0121407.

[22] Lai, A. K., Biewener, A. A., & Wakeling, J. M. (2018). Metabolic cost underlies task-dependent variations in motor unit recruitment. Journal of the Royal Society Interface, 15(148), 20180541.

[23] Song, S., & Geyer, H. (2015). A neural circuitry that emphasizes spinal feedback generates diverse behaviours of human locomotion. The Journal of physiology, 593(16), 3493-3511.

[24] Wang, J. M., Hamner, S. R., Delp, S. L., & Koltun, V. (2012). Optimizing locomotion controllers using biologically-based actuators and objectives. ACM Transactions on Graphics (TOG), 31(4), 1-11.

[25] Crowninshield, R. D., & Brand, R. A. (1981). A physiologically based criterion of muscle force prediction in locomotion. Journal of biomechanics, 14(11), 793-801.

[26] Grosse-Lordemann, H., & Müller, E. A. (1936). Der Einfluß der Leistung und der Arbeitsgeschwindigkeit auf das Arbeitsmaximum und den Wirkungsgrad beim Radfahren. Arbeitsphysiologie, 9(4), 454-475.

[27] Nguyen, V. Q., Johnson, R. T., Sup, F. C., & Umberger, B. R. (2019). Bilevel optimization for cost function determination in dynamic simulation of human gait. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(7), 1426-1435.

[28] Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2013). Principles of Neural Science (5P th P Edition).

[29] Menz, H. B., Lord, S. R., & Fitzpatrick, R. C. (2003). Acceleration patterns of the head and pelvis when walking on level and irregular surfaces. Gait & posture, 18(1), 35-46.

[30] Lin, D. C., McGowan, C. P., Blum, K. P., & Ting, L. H. (2019). Yank: the time derivative of force is an important biomechanical variable in sensorimotor systems. Journal of experimental biology, 222(18), jeb180414.

[31] Kavounoudias, A., Roll, R., & Roll, J. P. (1998). The plantar sole is a ‘dynamometric map’ for human balance control. Neuroreport, 9(14), 3247-3252.

[32] Kennedy, P. M., & Inglis, J. T. (2002). Distribution and behaviour of glabrous cutaneous receptors in the human foot sole. The Journal of physiology, 538(3), 995-1002.

[33] Mündermann, A., Dyrby, C. O., & Andriacchi, T. P. (2005). Secondary gait changes in patients with medial compartment knee osteoarthritis: increased load at the ankle, knee, and hip during walking. Arthritis & rheumatism, 52(9), 2835-2844.

[34] Nigg, B. M. (1985). Biomechanics, load analysis and sports injuries in the lower extremities. Sports medicine, 2(5), 367-379.
[35] Winter, D. A. (1991). Biomechanics and motor control of human gait: normal, elderly and pathological.
[36] Sloot, L. H., Van der Krogt, M. M., & Harlaar, J. (2014). Self-paced versus fixed speed treadmill walking. Gait & posture, 39(1), 478-484.
[37] Geijtenbeek, T. (2019). Scone: Open source software for predictive simulation of biological motion. Journal of Open Source Software, 4(38), 1421.
[38] Hansen, N. (2006). The CMA evolution strategy: a comparing review. Towards a new evolutionary computation: Advances in the estimation of distribution algorithms, 75-102.
[39] Hunt, K. H., & Crossley, F. R. E. (1975). Coefficient of restitution interpreted as damping in vibroimpact.
[40] Geyer H, Herr H. A muscle-reflex model that walks and runs. IEEE Trans Neural Syst Rehabil Eng. 2010;18(3):263–273.
[41] Umberger, B. R., Gerritsen, K. G., & Martin, P. E. (2003). A model of human muscle energy expenditure. Computer methods in biomechanics and biomedical engineering, 6(2), 99-111.
[42] Uchida, T. K., Hicks, J. L., Dembia, C. L., & Delp, S. L. (2016). Stretching your energetic budget: how tendon compliance affects the metabolic cost of running. PloS one, 11(3), e0150378.
[43] Van Criekinge, T., Saeys, W., Truijen, S., Vereeck, L., Sloot, L. H., & Hallemans, A. (2023). A full-body motion capture gait dataset of 138 able-bodied adults across the life span and 50 stroke survivors. Scientific data, 10(1), 852.
[44] Weinhandl, J. T., & Bennett, H. J. (2019). Musculoskeletal model choice influences hip joint load estimations during gait. Journal of biomechanics, 91, 124-132.
[45] Karatsidis, A., Bellusci, G., Schepers, H. M., De Zee, M., Andersen, M. S., & Veltink, P. H. (2016). Estimation of ground reaction forces and moments during gait using only inertial motion capture. Sensors, 17(1), 75.
[46] Dorschky, E., Nitschke, M., Seifer, A. K., Van Den Bogert, A. J., & Eskofier, B. M. (2019). Estimation of gait kinematics and kinetics from inertial sensor data using optimal control of musculoskeletal models. Journal of biomechanics, 95, 109278.
[47] Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). Partial least squares structural equation modeling (PLS-SEM) using R: A workbook (p. 197). Springer Nature.
[48] Bovi, G., Rabuffetti, M., Mazzoleni, P., & Ferrarin, M. (2011). A multiple-task gait analysis approach: kinematic, kinetic and EMG reference data for healthy young and adult subjects. Gait & posture, 33(1), 6-13.
[49] Devaprakash, D., Weir, G. J., Dunne, J. J., Alderson, J. A., & Donnelly, C. J. (2016). The influence of digital filter type, amplitude normalisation method, and co-contraction algorithm on clinically relevant surface electromyography data during clinical movement assessments. Journal of Electromyography and Kinesiology, 31, 126-135.
[50] Millard, M., Uchida, T., Seth, A., & Delp, S. L. (2013). Flexing computational muscle: modeling and simulation of musculotendon dynamics. Journal of biomechanical engineering, 135(2), 021005.

Appendix A

Table A1. Overview of the muscle parameters in the musculoskeletal model.

MTU in model	Maximum isometric force	Optimal fiber length	Tendon slack length	Pennation angle
Hamstrings	2594	0.0976	0.319	0.2025
Biceps femoris short head	804	0.1103	0.095	0.2147
Gluteus maximus	1944	0.1569	0.111	0.3822
Iliopsoas	2186	0.1066	0.152	0.2496
Rectus femoris	1169	0.0759	0.3449	0.2426
Vasti	4530	0.0993	0.1231	0.0785
Gastrocnemius	2241	0.051	0.384	0.1728
Soleus	3549	0.044	0.248	0.4939
Tibialis anterior	1579	0.0683	0.243	0.1676
Gluteus medius	2045	0.0733	0.066	0.3578
Adductor magnus	2268	0.087	0.06	0.0873

Table A2. Overview of gait phases and muscle reflexes used in the controller

	Early Stance	Late Stance	Lift Off	Swing	Landing
Hamstrings	L+, V+, F+ from gastroc, P from pelvis tilt	L+, V+, F+ from gastroc, P from pelvis tilt	L, V+,	L, V+,	L, V+,
Biceps femoris short head	L+, V+, F+ from gastroc	L+, V+, F+ from gastroc	F+	F+	F+
Gluteus maximus	L+	L+	L	L	L
Iliopsoas	2186	0.1066	L+, L from hamstrings, P from pelvis tilt	L+, L from hamstrings, P from pelvis tilt	L+, L from hamstrings, P from pelvis tilt
Rectus femoris	L+, P from pelvis tilt	L+, P from pelvis tilt	L+	L+	L+
Vasti	F+, L+, V+ L-&V- from bifemsh	F+, L+, V+ L-&V- from bifemsh			
Gastrocnemius	F, L+, F- from tib ant	F, L+, F- from tib ant	F, L+, F- from tib ant	F, L+, F- from tib ant	F, L+, F- from tib ant
Soleus	F, L+, F- from tib ant	F, L+, F- from tib ant	F, L+, F- from tib ant	F, L+, F- from tib ant	F, L+, F- from tib ant
Tibialis anterior	F, L+, F-&L- from soleus, F-&L- from gastroc	F, L+, F-&L- from soleus, F-&L- from gastroc	F, L+, F-&L- from soleus, F-&L- from gastroc	F, L+, F-&L- from soleus, F-&L- from gastroc	F, L+, F-&L- from soleus, F-&L- from gastroc
Gluteus medius	L+, V+, P from pelvis list	L+, V+, P from pelvis list	PD from hip adduction	PD from hip adduction	PD from hip adduction
Adductor magnus	L+, V+, P from pelvis list	L+, V+, P from pelvis list	PD from hip adduction	PD from hip adduction	PD from hip adduction

Appendix B

Figure B1. RMSE/SD between predictive results obtained with the optimal combination and ExpTrack and the experimental data.

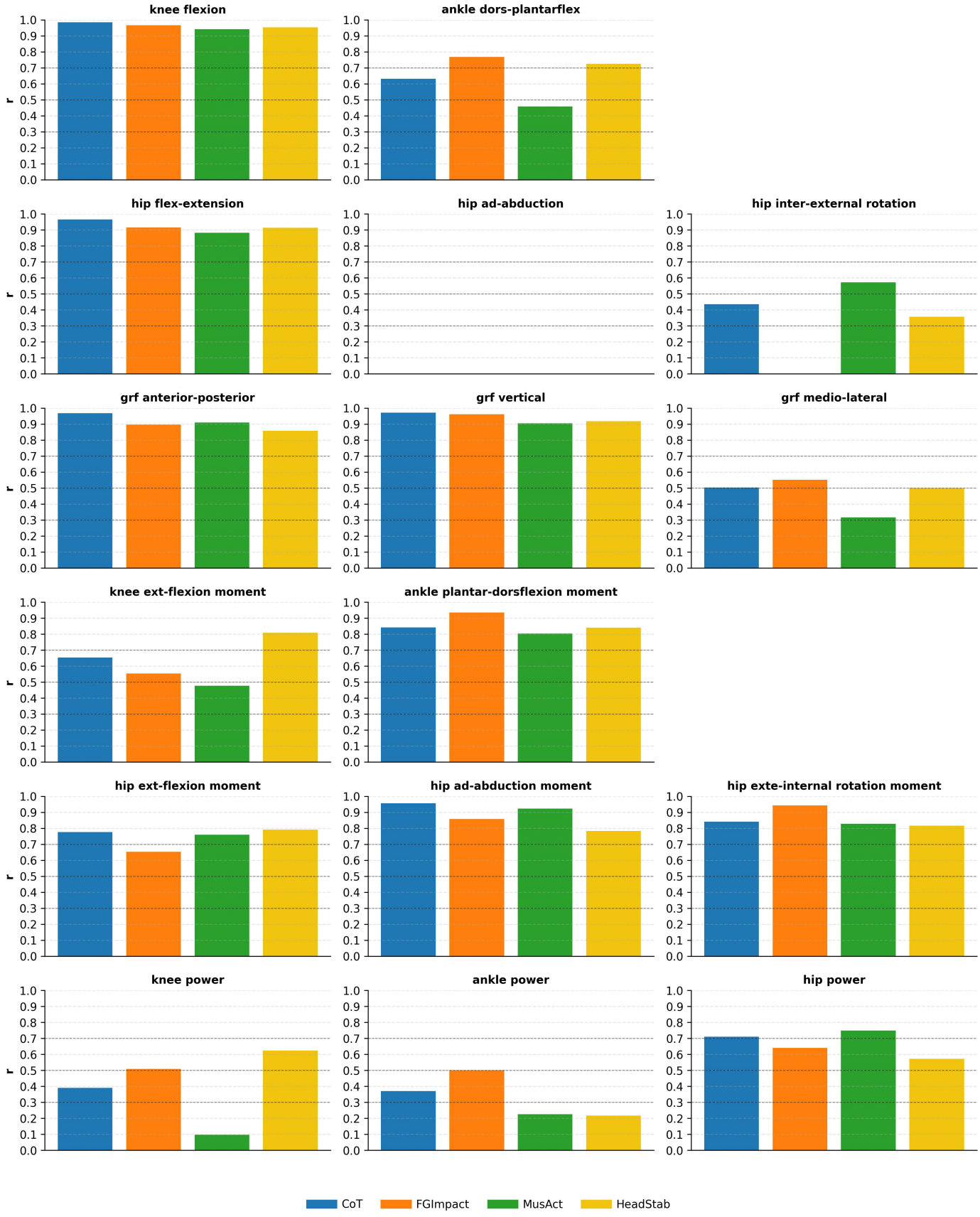


Figure B2. Pearson r between results by minimising four criteria and the experimental data. For hip ad-abduction and hip internal-external rotation, the absence of bars indicates negative r values and therefore very poor agreement.

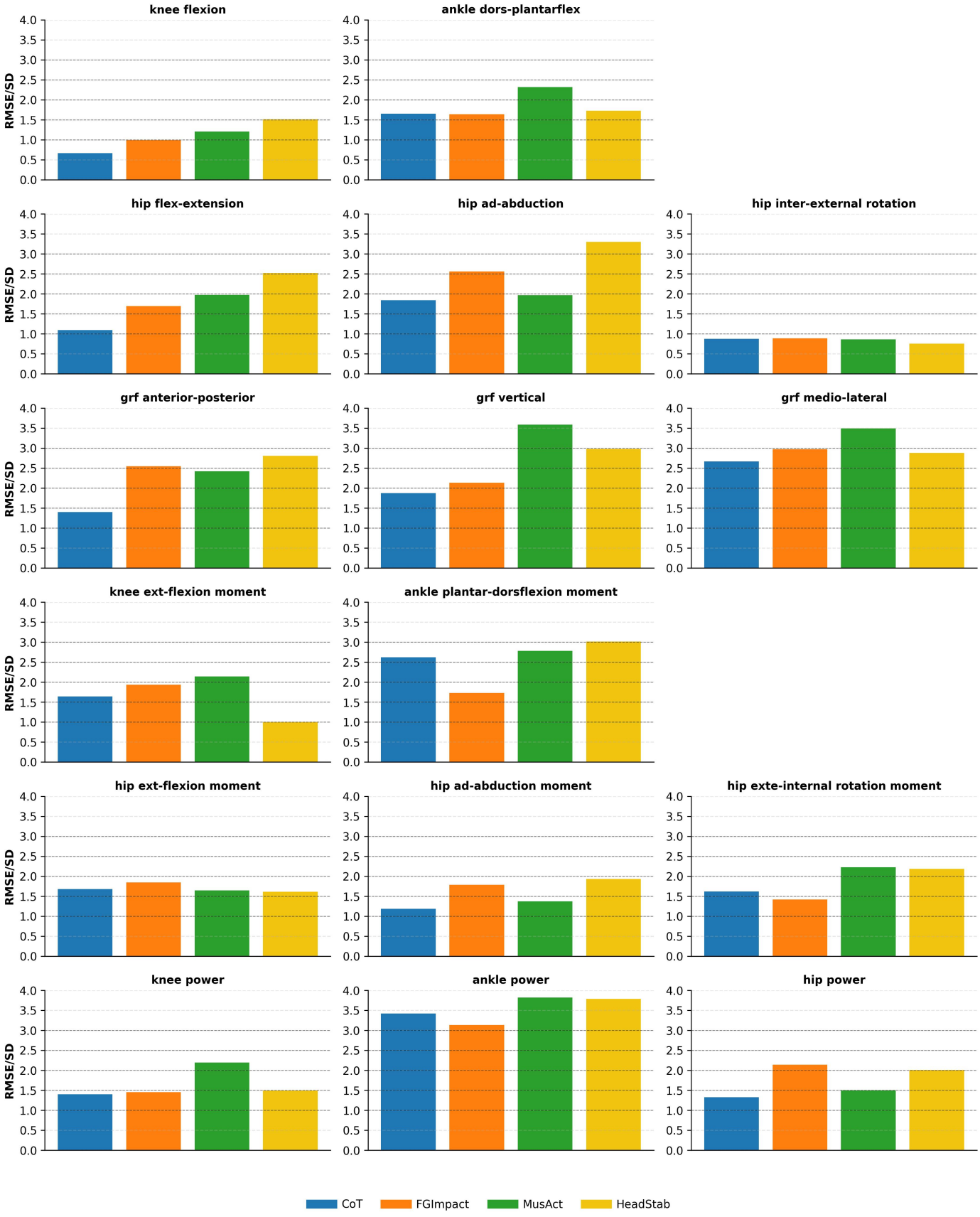


Figure B3. RMSE/SD between results by minimising four criteria and the experimental data.

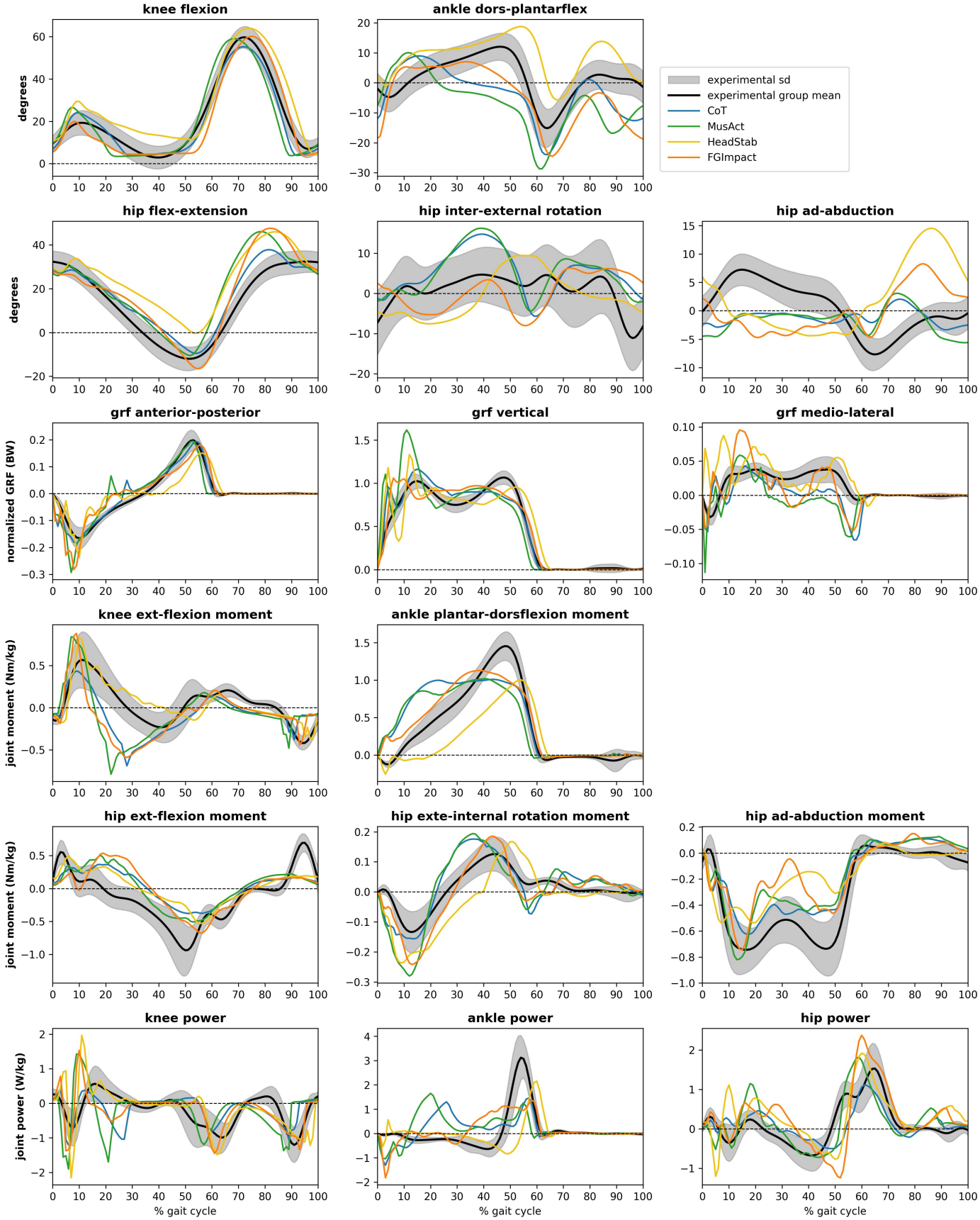


Figure B4. Comparison between gait patterns predicted by minimising four criteria and the experimental data.

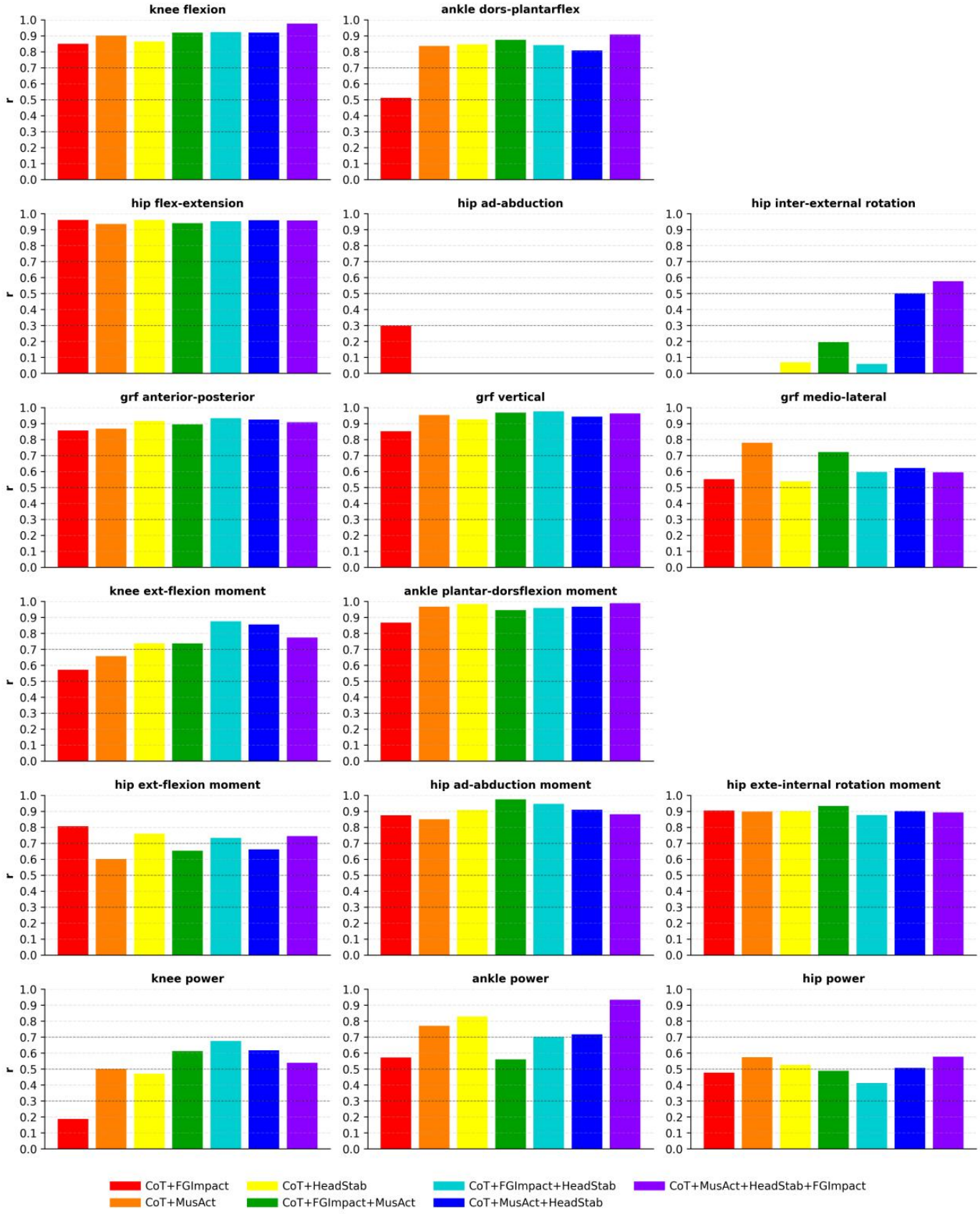


Figure B5. Pearson r between results by minimising different objective function combinations and the experimental data. For hip ad-abduction and hip internal-external rotation, the absence of bars indicates negative r values and therefore very poor agreement.

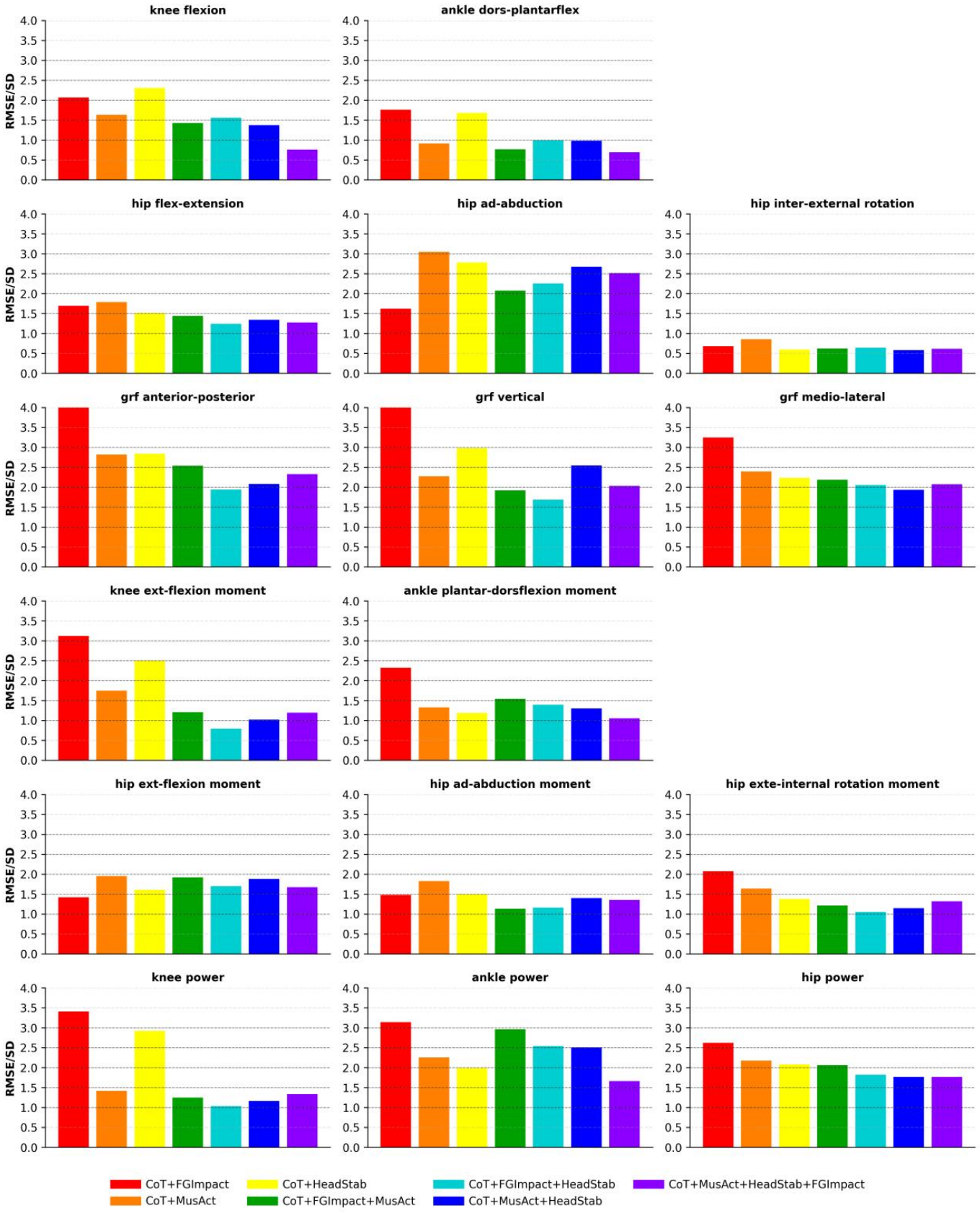


Figure B6. RMSE/SD between results by minimising different objective function combinations and the experimental data.

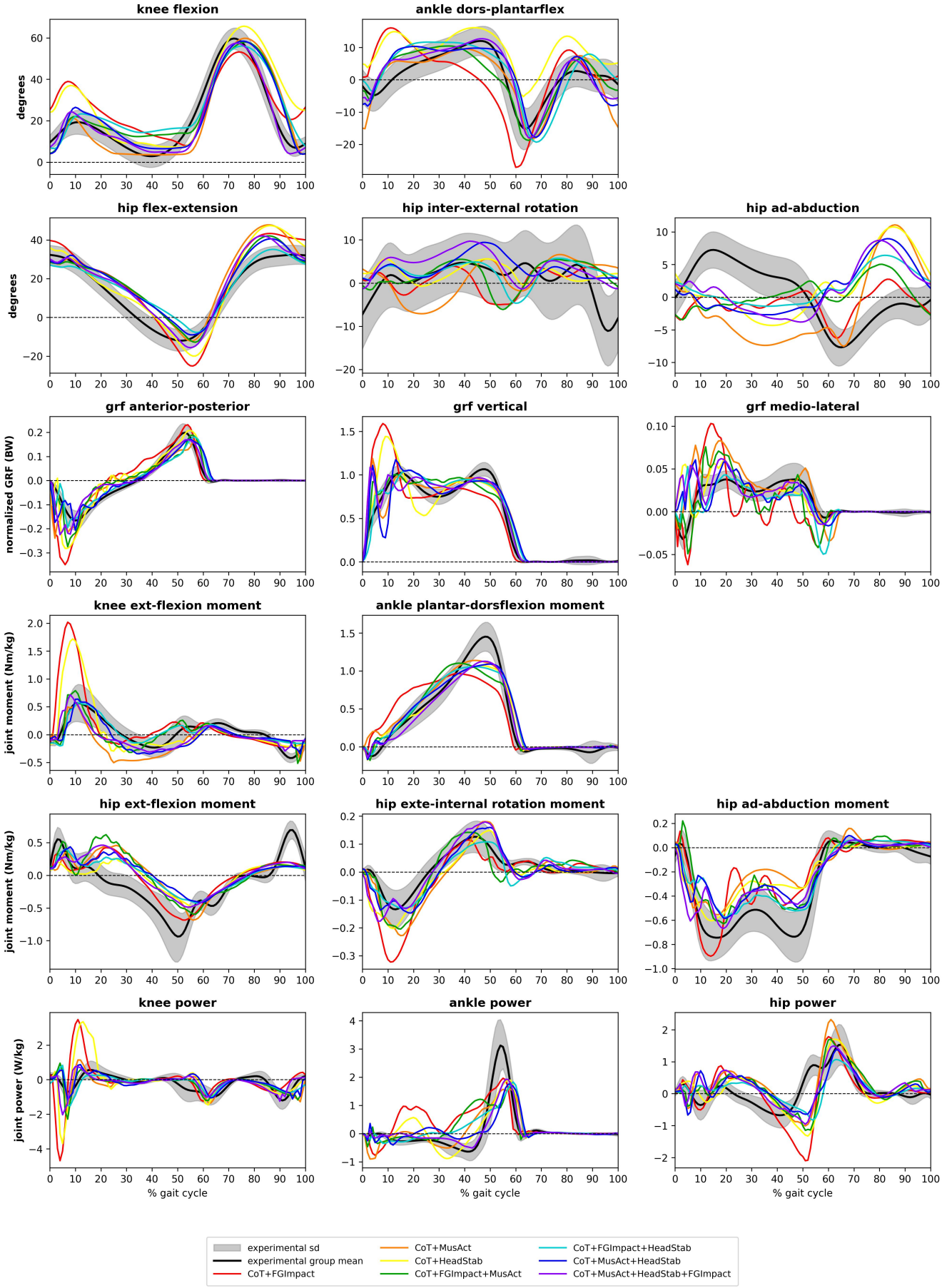


Figure B7. Comparison between gait patterns predicted by minimising different objective function combinations and the experimental data.