Nature Communications

https://doi.org/10.1038/s41467-026-75981-y

Article in Press

Biomechanics-informed inertial tracking achieves the accuracy of marker-based kinematics

Received: 14 December 2025

Accepted: 2 July 2026

Published online: 05 August 2026

Cite this article as: Phan, V., Li, Z., Meinders, E. *et al.* Biomechanics-informed inertial tracking achieves the accuracy of marker-based kinematics. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-75981-y

Vu Phan, Zhixiong Li, Evy Meinders, Tom Gale, William Anderst, Julian Ng-Thow-Hing, Aminreza Khandan & Eni Halilaj

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

aBiomechanics-Informed Inertial Tracking Achieves the Accuracy of Marker-Based Kinematics

Vu Phan[1],

Zhixiong Li[1],

Evy Meinders[1],

Tom Gale[2],

William Anderst[2],

Julian Ng-Thow-Hing[3],

Aminreza Khandan[1],

Eni Halilaj[1,2,4,5] ✉

[1]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA

[2]Department of Orthopaedic Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA

[3]Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA

[4]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA

[5]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA

✉ Please direct correspondence to:

Eni Halilaj, PhD

5000 Forbes Ave, Pittsburgh, PA, USA 15213

Email: ehalilaj@andrew.cmu.edu | Phone: 412-268-2183

ARTICLE IN PRESS

ARTICLE IN PRESS

ABSTRACT

Inertial measurement units (IMUs) could transform human movement science by enabling motion tracking outside the laboratory. Yet, their accuracy is perceived as inferior to that of marker-based systems. Here, we introduce IMoveLab, a method that integrates biomechanical priors into state-estimation filters, harnessing coupling across degrees of freedom and the spring-like behavior of muscle–tendon units for recursive state correction. We found that modern state-estimation filters match the accuracy of marker-based knee tracking when evaluated against biplane radiography, the gold standard for skeletal motion. IMoveLab further improves accuracy and achieves drift-free lower-extremity kinematics in long-duration trials, whereas alternative approaches do not. Movement scientists have been anticipating the leap out of the laboratory with cautious optimism. With methods and evidence that support more confident adoption of IMUs, this leap seems less like an aspiration and more like the first line of a new chapter written in the environments where human movement unfolds.

ARTICLE IN PRESS

ARTICLE IN PRESS

INTRODUCTION

In the past decade, inertial measurement units (IMUs) have become a pivotal tool in the field of human movement biomechanics, promising to galvanize research, clinical practice, and motor-skill acquisition at large. What sets IMUs apart from other tools is not only the promise of a more realistic view into human movement, which may be inhibited in the traditional laboratory environment[1–3], but also the fact that the knowledge acquired in research studies is easily translatable. The insights derived from traditional marker-based motion capture have been valuable, but often difficult to act upon. For example, although reduced knee flexion and peak joint loads following anterior cruciate ligament reconstruction have been related to post-traumatic osteoarthritis in traditional studies[4,5], typical clinical centers lack the laboratories required to measure these biomechanical outcomes and use them to personalize rehabilitation. IMUs are therefore one of the few emerging tools that researchers, clinicians, and athletic trainers are welcoming with equal enthusiasm and hopes for a stronger feedback loop between research and practice. Practitioners having access to research-grade data and researchers having access to clinical- or consumer-scale data is now a realistic future.

Adoption and scaling, however, have been stifled by the wait for established standards around the use of IMUs, accompanied by more active discussions on accuracy targets and syntheses of how different data-collection procedures and computational tools compare toward meeting those targets. Advances in microelectromechanical systems have made IMUs more miniaturized and power efficient. In addition to hard-cased sensors, elegant sticker-like ones that comply with the skin, resembling bandages, are now available to encourage user compliance[6]. Infrastructures for data transmission through protected cloud services and digital platforms for real-time data processing and ingestion into the electronic health record are also being developed. However, state estimation with IMUs is still met with reluctance in human-movement applications. Both commercial and open-source solutions are being actively developed, but the field currently lacks an understanding of how these methods compare because each has been tested on different datasets that are not publicly available. Validation studies often focus on specific, sometimes black-box, systems, while most reviews[7–9] have not classified findings by computational approach. Prior benchmarking efforts have stopped at estimating the orientation of single body segments with state-

ARTICLE IN PRESS

estimation filters[10,11], but in many biomechanics applications, joint kinematics expressed using the Grood and Suntay coordinate-system convention[12,13] are more informative. Due to ongoing knowledge gaps, the recent recommendations from the International Society of Biomechanics[14] still left the choice of computational approaches open-ended and non-prescriptive. Adding to these uncertainties, marker-based motion tracking is widely used as a reference in validation studies because it is the most common laboratory tool for motion tracking. Efforts to train machine learning models that learn from marker-based kinematics are therefore not uncommon[15–17]. The assumption has been that even though marker-based motion capture is not an ideal ground truth, it is more accurate than inertial tracking. This assumption remains to be tested.

State-estimation filters[18–20] remain the foundation of IMU-based kinematics estimation and have a long track record of success in navigation, aviation, and robotics, where they estimate the motion of rigid bodies. Complementary filters like Madgwick (MAD)[18] and Mahony (MAH)[19] combine short-term estimates from gyroscope data with long-term estimates from accelerometer, and sometimes magnetometer, data to provide computationally efficient estimation, typically for real-time embedded applications. Kalman filters and their nonlinear extensions, such as Extended Kalman Filters (EKF)[21–23], model system dynamics probabilistically and iteratively correct estimates based on sensor measurements to provide more accurate estimation, especially in the presence of uncertainty. Sharing characteristics from the two approaches, the Versatile Quaternion-based Filter (VQF)[10,24] has recently emerged as an adaptive filter for accurate state estimation. Its design explicitly incorporates gyroscope bias estimation, filtering of acceleration measurements, and rejection of magnetic disturbances, which may be advantageous in human-movement applications where soft-tissue motion and environmental perturbations may violate standard filter assumptions. The Robust IMU-based Attitude Neural Network (RIANN), on the other hand, is a data-driven filter that leverages deep neural networks[17]. With the abundance of state-estimation tools and their success in other fields, why is inertial tracking still a grand challenge in biomechanics?

Human movement is far more intricate to track because soft-tissue motion causes IMUs to move relative to the underlying bones, and the frequency content of soft-tissue motion overlaps significantly with that of human movement. To reduce the effect of soft-tissue artifact, a few systems apply inverse kinematics

ARTICLE IN PRESS

constrained by a biomechanical model after a state-estimation filter[14,25–27] (e.g., OpenSense[25]). Whether this additional step improves accuracy or mitigates drift, however, remains unknown. Claims of drift in state-estimation filters are common, but no studies to date have evaluated drift in trials longer than 10 minutes. In such long-duration scenarios, it is thought that inverse kinematics via a biomechanical model could be especially helpful because of the model's constraints[25]. Despite these hypothetical benefits, we identified two key limitations arising from the sequential application of biomechanical modeling after a state-estimation filter. This approach, inspired by marker-based inverse kinematics, does not consider the Markovian property of IMU-based state estimation[28], applying box constraints that only cap kinematics at range-of-motion limits. However, much can go awry in state estimation before physiological limits are violated.

In this study, we propose an approach to integrate biomechanical constraints into the recursive process of state-estimation filters (Figs. 1 and 2). Our approach, IMoveLab, focuses on lower-extremity joints and uses coupling in the knee joint and muscle-tendon-unit constraints in the hip and ankle joints to achieve accurate and drift-free kinematics. Because these constraints are applied recursively at each time step, kinematic error is kept from amplifying over time. Before comparing our approach against existing ones, we first investigated whether marker-based analysis should continue to be used as a reference in studies seeking to improve inertial-tracking accuracy. Using biplane radiography[29] as the gold standard for skeletal motion (Fig. 1), we tested what we consider one of the most anticipated hypotheses of the past decade in the field of human motion tracking: that marker-based motion tracking is indeed more accurate than inertial tracking, focusing specifically on the knee (H1). We then hypothesized that our proposed approach, IMoveLab, would improve accuracy (H2.A) and reduce drift (H2.B) compared to both baseline state-estimation filters and OpenSense. We assessed the accuracy of knee kinematics using biplane radiography as the reference, during four activities (H2.A; Fig. 1). Drift was assessed across the three lower-extremity joints, using marker-based analysis as a reference, in a short-duration protocol that included eleven tasks (H2.B1; Fig. 2) and a long-duration one focusing on walking (H2.B2; Fig. 2). The combination of short-duration single-joint biplane radiography, short-duration multi-joint and multi-task marker-based analysis, and long-duration

multi-joint marker-based analysis enabled a principled balance between experimental rigor and practical feasibility.

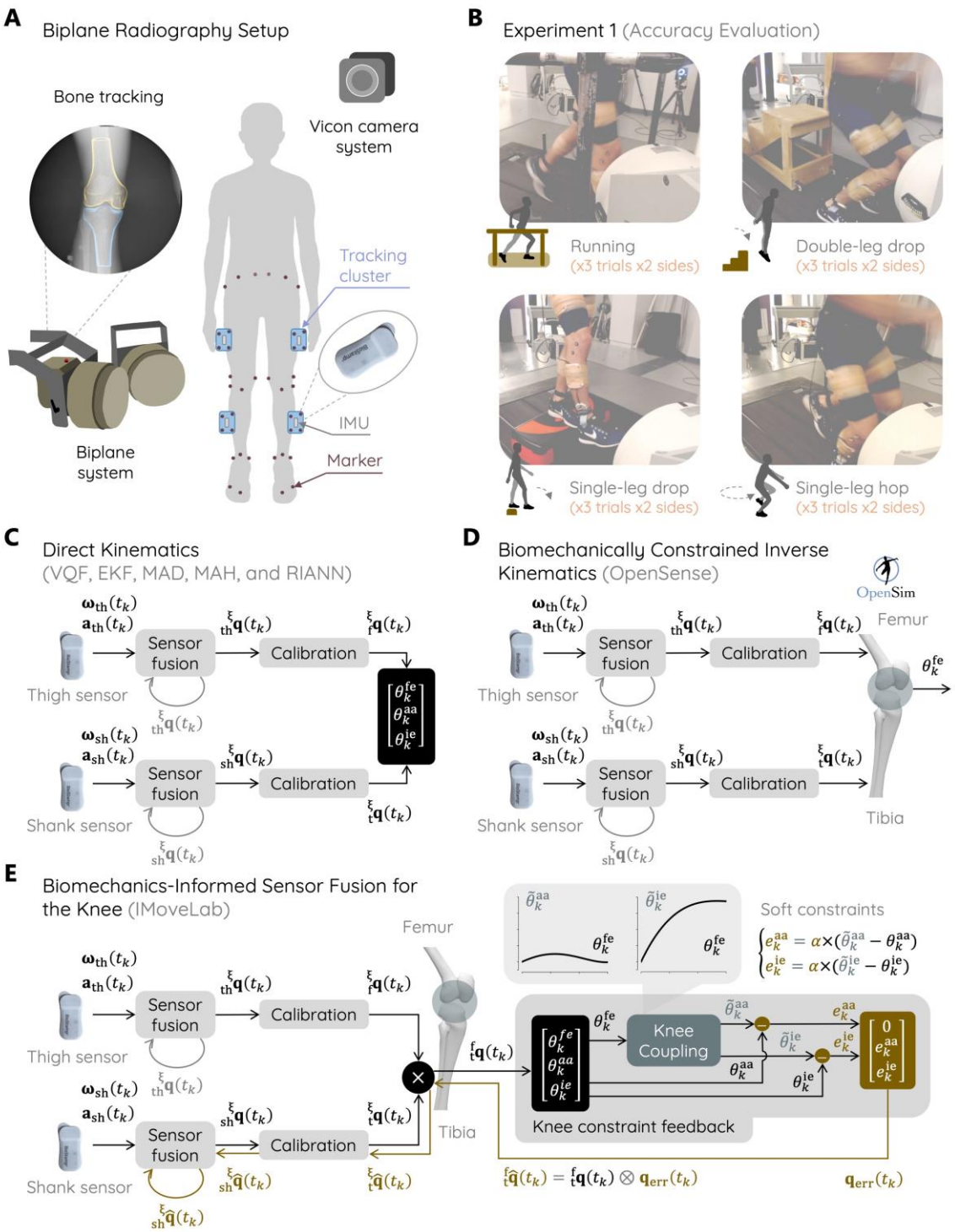

Figure 1. Overview of Experiment 1 (with Ground-Truth Biplane Radiography for Skeletal Kinematics). A) Participants were instrumented with four inertial measurement units (IMUs) directly attached to the skin of their thighs

ARTICLE IN PRESS

and shanks with double-sided adhesive tape, reflective markers, and tracking clusters wrapped around body segments with self-adherent elastic bandage. A biplane-radiography system was used to obtain ground-truth knee kinematics. **B)** Each participant performed running and three functional exercises, three times per leg. **C)** For direct kinematics, sensor orientations were obtained by fusing inertial data using baseline filters. Sensor-to-body calibration was carried out using functional calibration. **D)** For biomechanically constrained inverse kinematics, we used OpenSense. **E)** Simultaneously, we applied our approach, IMoveLab, a biomechanics-informed sensor-fusion framework in which biomechanical constraints are embedded within the recursive state-estimation process, instead of being applied post hoc. At the knee, these constraints encode coupling across degrees of freedom. In (C)-(E), $\boldsymbol{\omega} \in \mathbb{R}^3$: three-axis gyroscope data, $\mathbf{a} \in \mathbb{R}^3$: three-axis accelerometry data, $\mathbf{q}, \hat{\mathbf{q}} \in \mathbb{R}^4$: orientation in quaternion notation, $\theta, \tilde{\theta} \in \mathbb{R}^1$: Euler angle, and $e \in \mathbb{R}^1$: Euler angle error. Left subscripts are used for the thigh (th), shank (sh), femur (f), and tibia (t) and right subscripts for the timestep (k) and error (err). A left superscript denotes the global reference frame of the inertial sensors (ξ), whereas right superscripts are used for flexion/extension (fe), adduction/abduction (aa), internal/external rotation (ie).

ARTICLE IN PRESS

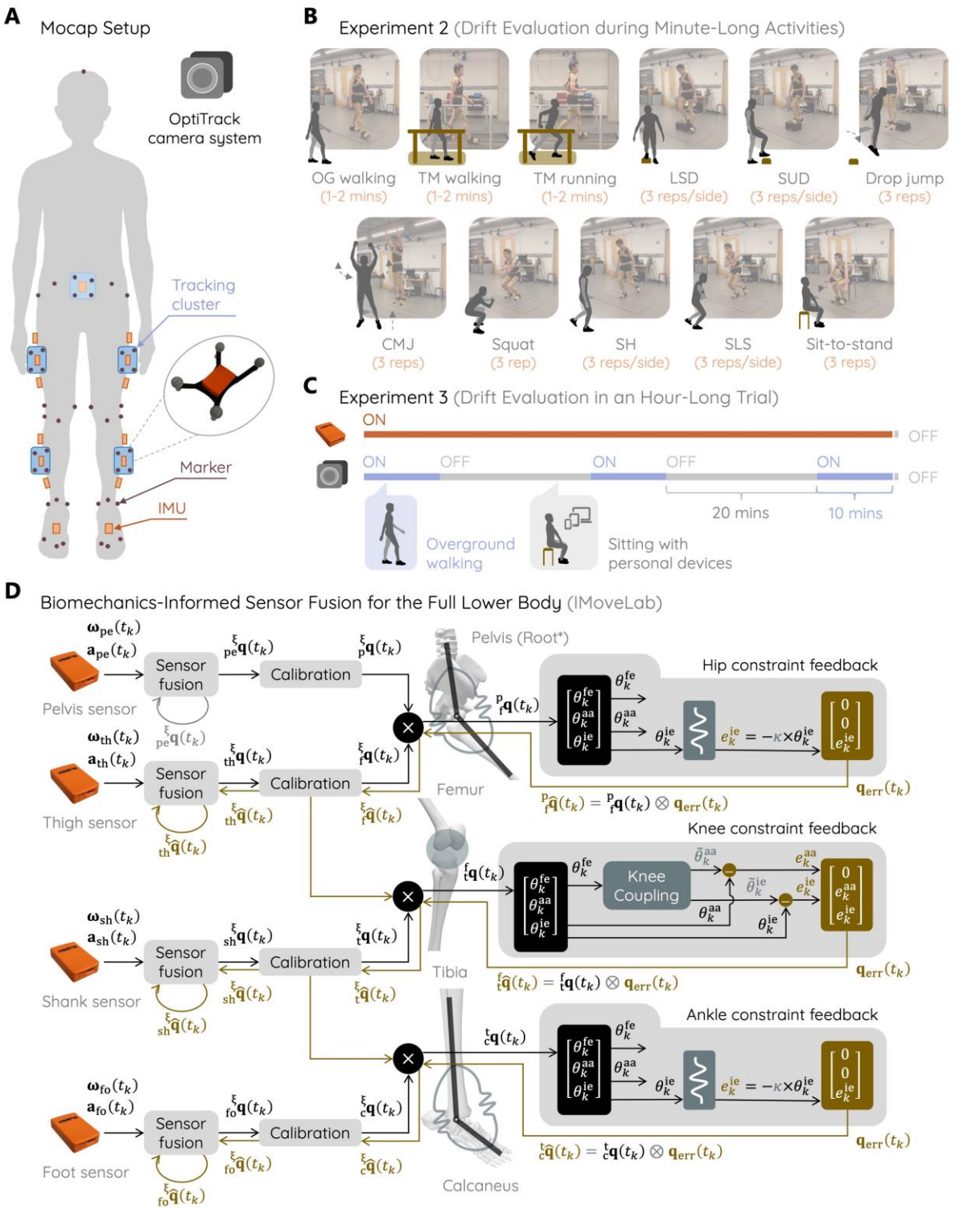

Figure 2. Overview of Experiments 2 and 3 (with Marker-Based Motion Capture as the Reference). A) Participants were instrumented with 15 inertial measurement units (IMUs) on the lower body: one on the lower back, one on each foot, and three on each thigh and shank, with the redundancy enabling evaluation of the effect of sensor placement.

ARTICLE IN PRESS

Marker-based motion capture was used to obtain reference kinematics. **B)** In Experiment 2, each participant performed three locomotion tasks and eight functional exercises. **C)** In Experiment 3, a 70-minute trial, with three sequences of 10-minute walking separated by 5 minutes of sitting and 15 minutes of idling, was added for a subset of the participants to study the effect of drift. **D)** Kinematics of the hip, knee, and ankle were obtained directly from state-estimation filters, using biomechanically constrained inverse kinematics (OpenSense), and using our biomechanics-informed sensor-fusion method (IMoveLab). Constraints at the hip and ankle encode the spring-like behavior of muscle–tendon units, introducing a restoring term that counteracts drift in kinematic trajectories by penalizing deviations from physiologically plausible kinematics. In the diagram, OG stands for overground, TM for treadmill, LSD for lateral step down, SUD for step up and down, CMJ counter-movement jump, SH for step and hold, and SLS for single-leg squat. In (C)-(D), $\boldsymbol{\omega} \in \mathbb{R}^3$: three-axis gyroscope data, $\mathbf{a} \in \mathbb{R}^3$: three-axis accelerometry data, $\mathbf{q}, \widehat{\mathbf{q}} \in \mathbb{R}^4$: orientation in quaternion notation, $\theta, \tilde{\theta} \in \mathbb{R}^1$: Euler angle, and $e \in \mathbb{R}^1$: Euler angle error. Sensor locations are indicated by left subscripts: pelvis (pe), thigh (th), shank (sh), and foot (fo). Body segments are also indicated by left subscripts: pelvis (p), femur (f), tibia (t), and calcaneus (c). The left superscript ξ represents the global reference frame of the inertial sensors. Right superscripts fe, aa, and iestand for flexion/extension, adduction/abduction, and internal/external rotation, respectively. Right subscript k represents the timestep, while err stands for error.

RESULTS

Overview of IMoveLab: A Biomechanics-Informed Sensor-Fusion Framework

Our approach, IMoveLab, embeds known biomechanical relationships into sensor-fusion filters to improve kinematics-estimation accuracy and reduce drift compared to baseline filters. Instead of applying hard joint-limit constraints sequentially after sensor fusion, as is commonly done today (e.g., OpenSense), IMoveLab leverages joint coupling at the knee and spring-like muscle–tendon behavior at the hip and ankle to inject biomechanics information recursively into the state-estimation loop (see Methods for a detailed description). This formulation allows the system to correct drift as it emerges, particularly when reliable magnetometer data to correct heading drift are unavailable, while preserving a computational efficiency that lends itself well to real-time applications (e.g., assistive-device control).

Is Marker-Based Tracking More Accurate than Inertial Tracking? A Biplane-Radiography Study of the Knee

ARTICLE IN PRESS

We first tested the hypothesis that marker-based motion tracking is more accurate than inertial tracking using data from Experiment 1, where both inertial and marker-based knee kinematics were compared against gold-standard skeletal kinematics from biplane radiography (Fig. 1A–B), using root-mean-square difference (RMSD) as the primary outcome. We focused on the knee as it is one of the most studied lower-extremity joints, with a high prevalence of osteoarthritis[30] and the most common site of joint replacement surgery[31]. Unconstrained, or direct, kinematics were computed using baseline six-dimensional (6D) state-estimation filters (Fig. 1C). We focused on filters that operate with gyroscope and accelerometer data (i.e., 6D filters), given their versatility and conduciveness to testing within the biplane radiography system, under which magnetometers are unreliable. Overall, knee flexion/extension was similarly accurate between marker-based analysis and model-based state-estimation filters (median RMSDs = 4.1° [3.6°, 4.7°] for marker-based vs. 4.2° [3.2°, 5.2°] for VQF, 4.1° [3.8°, 5.1°] for EKF, 4.0° [3.5°, 4.3°] for MAD, and 5.7° [3.9°, 6.3°] for MAH, $p > 0.191$ for all comparisons). RIANN, the data-driven method, was significantly less accurate (RMSD = 6.2° [6.0°, 7.4°], $p < 0.001$) (Fig. 3A). Neither marker-based analysis nor inertial tracking was able to capture the waveform of knee adduction/abduction and internal/external rotation, despite median RMSDs being close to 5° (Fig. 3B–C, Supplementary Fig. 1). As a result, we did not perform statistical analyses on these secondary degrees of freedom.

When analyzed by activity, IMU-based kinematics obtained from the top three filters (i.e., VQF, EKF, and MAD) achieved knee flexion/extension accuracy comparable to that of marker-based analysis (Fig. 3D–G). Specifically, for running, filters showed no significant difference from the marker-based approach (RMSDs = 4.0° [3.5°, 5.0°] for marker-based vs. 4.7° [2.9°, 6.0°] for VQF, 4.1° [3.2°, 5.8°] for EKF, and 4.1° [3.3°, 5.7°] for MAD, $p > 0.542$ for all comparisons). A similar trend was observed in jumping tasks, such as single-leg hop (RMSDs = 4.2° [3.4°, 5.4°] for marker-based vs. 3.1° [2.2°, 4.1°] for VQF, 4.1° [3.1°, 5.7°] for EKF, and 3.0° [2.3°, 4.1°] for MAD, $p > 0.080$), single-leg drop (RMSD of 3.0° [2.7°, 3.7°] for marker-based vs. 3.6° [3.1°, 4.2°] for VQF, 3.5° [3.1°, 4.2°] for EKF, and 3.4° [2.9°, 3.7°] for MAD, $p > 0.080$), and double-leg drop (RMSDs = 4.7° [2.7°, 5.7°] for marker-based vs. 4.1° [3.0°, 6.4°] for VQF, 4.3° [3.3°, 6.2°] for EKF, and 4.1° [3.3°, 6.6°] for MAD, $p > 0.542$). When filters were as accurate as marker-based analysis, no significant relationship was observed between filter agreement with biplane radiography and marker-based analysis

(e.g., ρ = 0.335, p = 0.263 for VQF). A strong relation was observed when they were less accurate (e.g., ρ = 0.692, p = 0.009 for RIANN) (Fig. 3H–L). When grouping estimates separately using a threshold of 4.1° (i.e., disagreement of marker-based and biplane radiography), the group with RMSDs of less than 4.1° showed no correlation (ρ = 0.365, p = 0.087), while that with RMSDs of more than 4.1° showed significant correlation (ρ = 0.627, p < 0.001) (Supplementary Fig. 2).

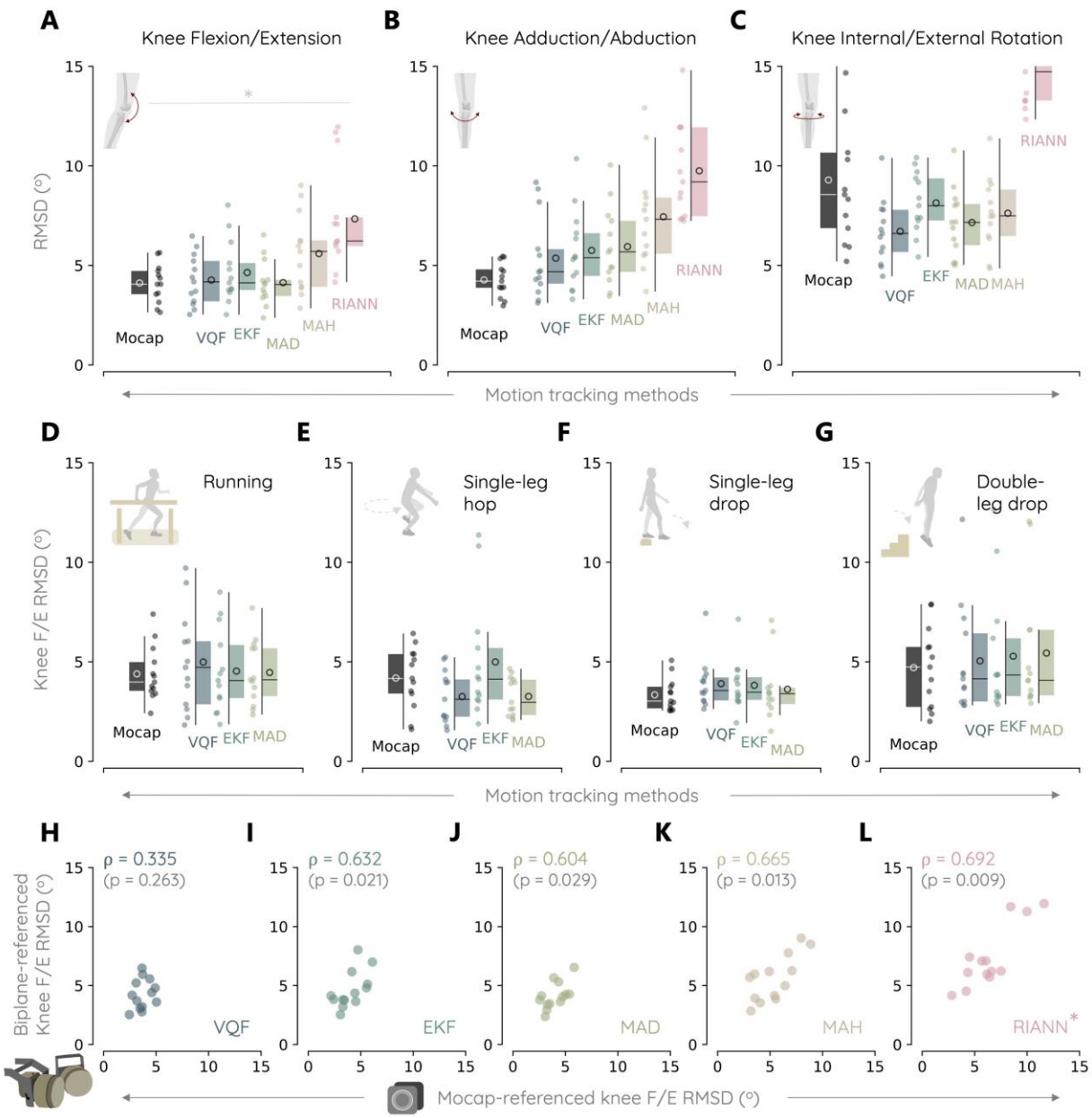

Figure 3. Comparisons of Inertial- and Marker-Based Motion Tracking with Ground-Truth Skeletal Kinematics from Biplane Radiography (H1 – Experiment 1). Each data point represents a participant (n = 13). **A)** Inertial- and

ARTICLE IN PRESS

marker-based kinematics exhibited similar accuracy in tracking knee flexion/extension, with the exception of the data-driven RIANN, which performed worse than the model-based filters. **B–C)** Since both inertial- and marker-based motion capture were not able to capture the waveforms of secondary degrees of freedom (Supplementary Fig. 1), no statistical analyses comparing these approaches against biplane radiography were performed. **D)** During running, **E)** single-leg hop, **F)** single-leg drop, and **G)** double-leg drop, the top three filters achieved similar accuracies to marker-based motion capture. Two-sided Wilcoxon signed-rank tests were used to test for differences between each filter (VQF, EKF, MAD) and marker-based motion capture (Mocap), with statistical significance (*) adjusted for multiple comparisons using a Holm-Bonferroni correction ($p < 0.003$). **H–L)** The flexion/extension (F/E) RMSDs when using marker-based motion capture (X axis) or biplane radiography (Y axis) as ground truth were correlated (two-sided Spearman tests) for filters with low accuracy (i.e., RIANN), but not ones with high accuracy (e.g., VQF). Statistical significance (*) was adjusted using a Holm-Bonferroni correction ($p < 0.013$). Supplementary Fig. 2 illustrates that the accuracy of the marker-based system sets the threshold for whether a relationship exists. The large vertical bars in (A)–(G) indicate the min-max range of the box plot, or Q1 – 1.5 * Interquartile range (IQR) to Q3 + 1.5 * IQR. Circles within the box plots denote mean, while horizontal bars denote median.

Do Biomechanical Constraints Improve Inertial-Tracking Accuracy?

Next, we tested hypothesis 2.A that our proposed approach IMoveLab (Fig. 1E) would achieve better accuracy compared to both the direct method using baseline state-estimation filters (Fig. 1C) and the biomechanically constrained inverse-kinematics method, OpenSense (Fig. 1D). To test this hypothesis, we used data from Experiment 1, with ground-truth knee kinematics from biplane radiography. We continued to focus on 6D filters, for reasons noted above. Inclusion of biomechanics significantly affected the accuracy of VQF, EKF, and RIANN ($p < 0.013$), but not that of MAD and MAH ($p > 0.058$) (Fig. 4A). Specifically, IMoveLab reduced knee flexion/extension RMSD from 4.2° [3.2°, 5.2°] to 2.8° [2.5°, 3.5°] for VQF ($p < 0.001$) and from 4.1° [3.8°, 5.1°] to 2.9° [2.6°, 3.8°] for EKF ($p = 0.002$), exhibiting lower errors than OpenSense, whose RMSD was 4.3° [3.3°, 5.7°] for VQF and 5.4° [4.1°, 8.6°] for EKF. The accuracy of OpenSense remained the same as that of direct kinematics for all the considered filters ($p > 0.057$), except for the data-driven RIANN, whose median RMSD was reduced from 6.2° [6.0°, 7.4°] to 5.1° [4.4°, 6.0°] ($p = 0.005$). IMoveLab could not be tested with RIANN because it is designed to correct estimates from model-based filters that accommodate feedback recursively, rather than end-to-end data-driven ones like RIANN.

ARTICLE IN PRESS

When estimating secondary degrees of freedom, IMoveLab significantly improved kinematics-estimation accuracy, reducing median RMSD to as low as 0.9° [0.8°, 1.0°] (with VQF) for knee adduction/abduction ($p < 0.001$ for all filters; Fig. 4B) and to as low as 3.7° [3.4°, 4.1°] (with MAH) for knee internal/external rotation ($p < 0.001$ for all filters; Fig. 4C). No comparison of secondary knee degrees of freedom was performed for OpenSense, as these outputs are not provided by the software. When the RMSD was evaluated against the marker-based reference, instead of biplane radiography, no significant differences were observed across the direct, OpenSense, and IMoveLab methods for any of the filters ($p > 0.018$, not significant after Holm-Bonferroni correction) (Fig. 4D). This lack of differences highlights the challenges of using marker-based analysis as a reference standard for accuracy assessment. Marker-based analysis may therefore be more appropriate for evaluation of drift than evaluation of accuracy moving forward.

ARTICLE IN PRESS

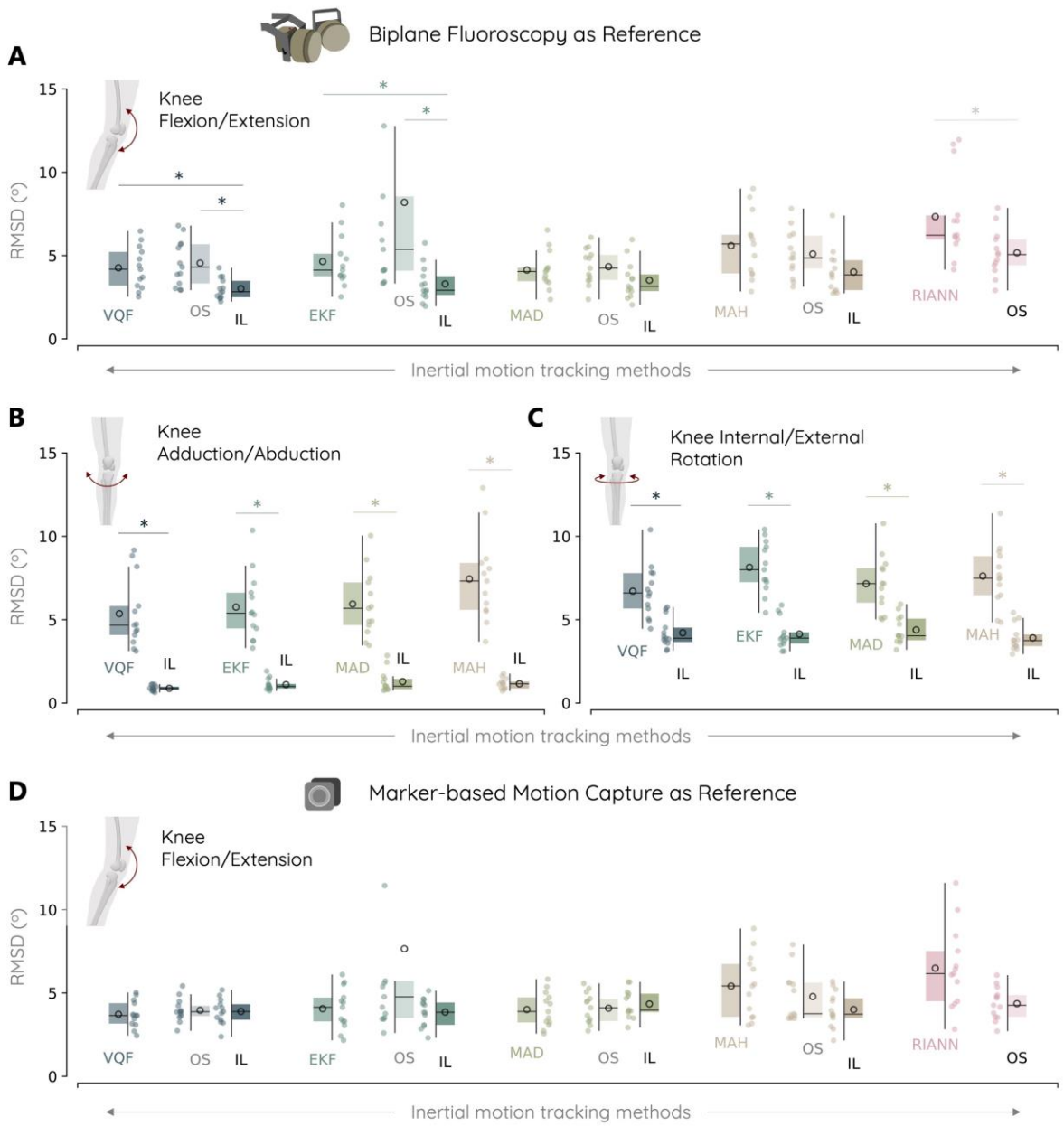

Figure 4. Impact of Biomechanical Constraints on Kinematics-Estimation Accuracy (H2.A – Experiment 1). Each data point represents a participant (n = 13). **A)** Our proposed approach, IMoveLab (IL), significantly improved the estimation of knee flexion/extension compared to the baseline VQF and EKF filters and compared to OpenSense (OS), when using the biplane radiography as a reference. OpenSense only improved the accuracy of RIANN but not that of the other filters. **B–C)** IMoveLab also reduced errors in knee adduction/abduction and knee internal/external rotation. **D)** However, when using marker-based kinematics as the reference, no significant differences across methods were observed. Friedman tests were used to evaluate the overall effect of tracking method (i.e., direct, IL, and OS) on estimation accuracy, followed by two-sided Wilcoxon signed-rank tests for post-hoc pairwise comparisons. The large

ARTICLE IN PRESS

vertical bars indicate the min-max range of the box plot, or Q1 – 1.5 * Interquartile range (IQR) to Q3 + 1.5 * IQR. Circles within the box plots denote mean, while horizontal bars denote median. Asterisks (*) indicate statistically significant differences for pairwise tests, after applying Holm-Bonferroni corrections (Wilcoxon $p < 0.05$). Pairwise comparisons were not carried out for MAD and MAH knee flexion/extension accuracy due to the lack of statistically significant differences in Friedman tests (Friedman $p < 0.025$).

Do Biomechanical Constraints Reduce Drift?

We then tested hypothesis 2.B, that our biomechanics-informed sensor-fusion approach, IMoveLab, would reduce drift better than the direct method with baseline state-estimation filters and better than OpenSense. To test this hypothesis, we used short-duration trials (i.e., 1 to 2 mins) from Experiment 2 (H2.B1) and a long-duration trial (i.e., 70 mins) from Experiment 3 (H2.B2) (Figs. 2A–C), with marker-based analysis as a reference. Given that these two experiments were conducted in a traditional gait lab, assessment of both 6D and 9D filters was feasible. Here, we focus on 6D filter results, with additional 9D results presented in Supplementary Figs. 3–5. Drift is defined as an increase in RMSD over time, and a drift-correction approach would attenuate this increase.

During 11 short activities lasting from one to two minutes (e.g., walking, squatting, jumping), both OpenSense and IMoveLab reduced drift over time overall across filters ($p < 0.006$), except for RIANN ($p = 0.819$), which did not experience drift (Fig. 5A) but generally had lower accuracy (Fig. 3A). However, the effect of IMoveLab was more salient, while that of OpenSense was less consistent. VQF experienced a small overall drift of 0.4° in these short-duration trials, and IMoveLab and OpenSense decreased this drift by similar amounts (0.3° [0.0°, 0.4°] vs. 0.2° [0.1°, 0.2°], $p = 0.123$). When broken down by degree of freedom, neither method reduced drift ($p > 0.016$, no significance after Holm-Bonferroni correction) (Fig. 5B). For EKF and MAD, which experienced more significant drift than VQF, IMoveLab reduced drift more than OpenSense across degrees of freedom (by 2.7° [2.3°, 3.2°] ($p < 0.001$) vs. 1.5° [0.7°, 2.2°], $p = 0.007$ for EKF; by 2.9° [2.6°, 3.9°] vs. 0.7° [0.4°, 1.2°], $p < 0.001$ for MAD) (Fig. 5A). For MAH, IMoveLab reduced drift by 6.9° [5.2°, 8.7°], while OpenSense increased it by 3.7° [1.0°, 6.2°] ($p < 0.001$). IMoveLab reduced drift in each degree of freedom ($p < 0.001$ for all comparisons) (Fig. 5C), whereas OpenSense improved

drift in most degrees of freedom ($p < 0.001$) except the knee flexion/extension ($p = 0.134$) (Fig. 5C). Qualitative illustrations of short-term drift can be found in Figs. 5D and E.

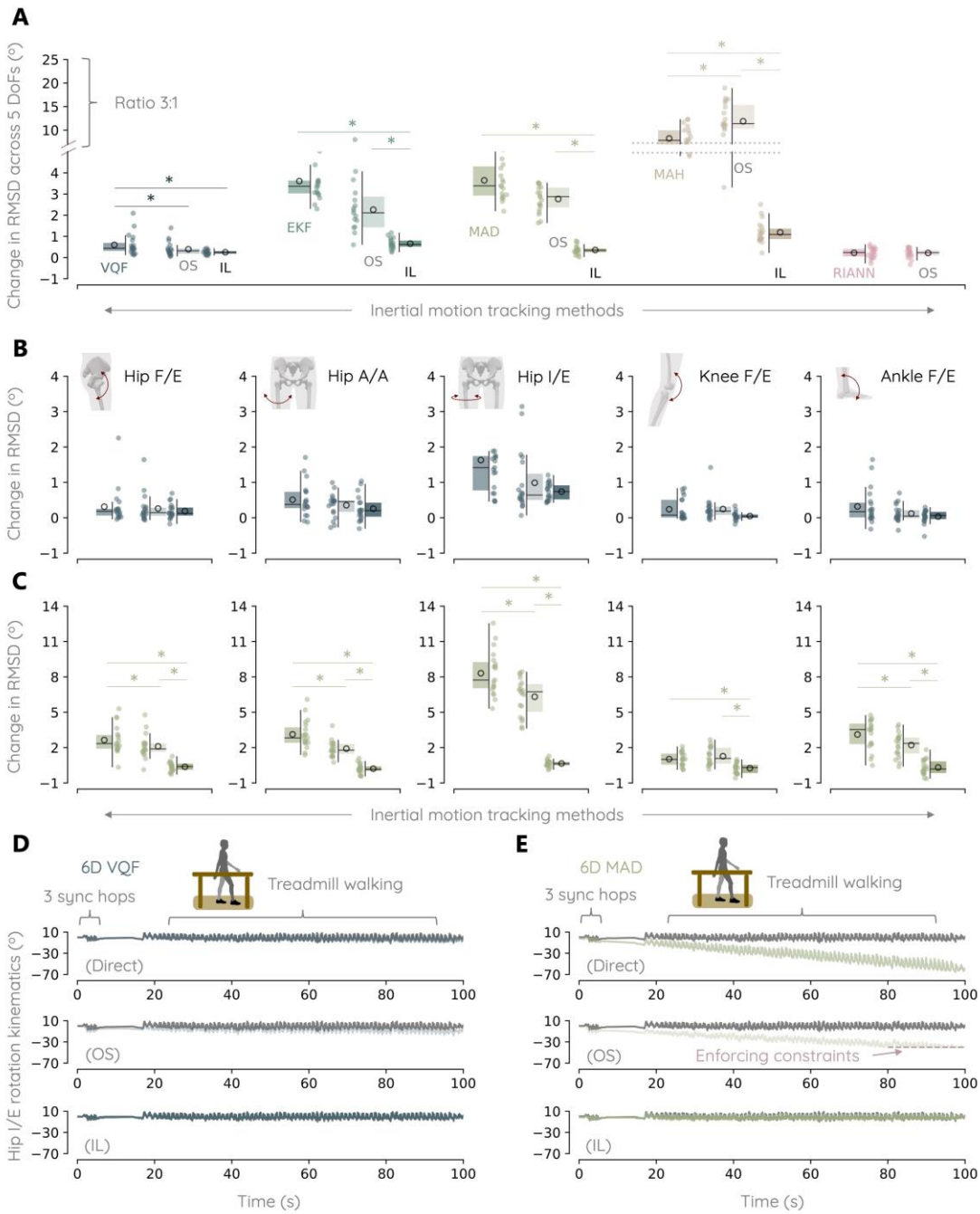

Figure 5. Impact of Biomechanical Modeling on Short-Term Drift (H2.B1 – Experiment 2). A) IMoveLab was more successful at reducing drift than OpenSense, especially for MAH, where it achieved significant drift reduction, while OpenSense worsened drift. RMSD is averaged across five degrees of freedom and 11 tasks. **(B)** VQF exhibited no drift in each degree of freedom in these short-duration trials. **(C)** In contrast, MAD, like other state-estimation filters, showed

ARTICLE IN PRESS

significant drift, which was mitigated by both OpenSense and IMoveLab. In (A)–(C), each data point represents a participant (n = 19). Friedman tests were used to evaluate the overall effect of tracking method (i.e., direct, IL, and OS) on drift, followed by two-sided Wilcoxon signed-rank tests for post-hoc pairwise comparisons. The large vertical bars indicate the min-max range of the box plot, or Q1 – 1.5 * Interquartile range (IQR) to Q3 + 1.5 * IQR. Circles within the box plots denote mean, while horizontal bars denote median. Asterisks (*) indicate statistically significant differences, after applying Holm-Bonferroni corrections (Wilcoxon $p < 0.05$). Kinematics from only five degrees of freedom were considered in this analysis, with the non-sagittal-plane kinematics of the knees and ankles excluded. F/E stands for flexion/extension, A/A for adduction/abduction, and I/E for internal/external rotation. Qualitative illustration of direct kinematics, OpenSense (OS), and IMoveLab, for **(D)** VQF and **(E)** MAD, shown here for one participant whose kinematics exhibited short-term drift with at least one of the two filters. Under drift, OS only enforces constraints at the joint limits, whereas IMoveLab enables immediate correction and drift-free kinematics during the entire trial.

In evaluating drift over the course of an hour, kinematics obtained directly from state-estimation filters and OpenSense suffered from significant drift, whereas those from IMoveLab achieved drift-free kinematics that were similar to those obtained with a 9D filter (Figs. 6A and B). Since 6D VQF and OpenSense suffered from significant drift within the first 10 minutes ($p < 0.001$), no changes in RMSD were observed across the three 10-minute walking trials captured at the beginning, middle, and end of the hour-long experiment. In contrast, IMoveLab achieved drift-free kinematics with an overall RMSD of less than 5° ($p = 0.091$) (Fig. 6C). Drift was also eliminated when magnetometer information was available (i.e., 9D VQF) (Fig. 6D). Qualitatively, all the methods had RMSDs of less than 5° within the first 2 minutes, but afterwards, the unconstrained 6D VQF and 6D VQF constrained with OpenSense drifted significantly. VQF was selected as a representative filter, but these results held true across filters (Supplementary Fig. 3).

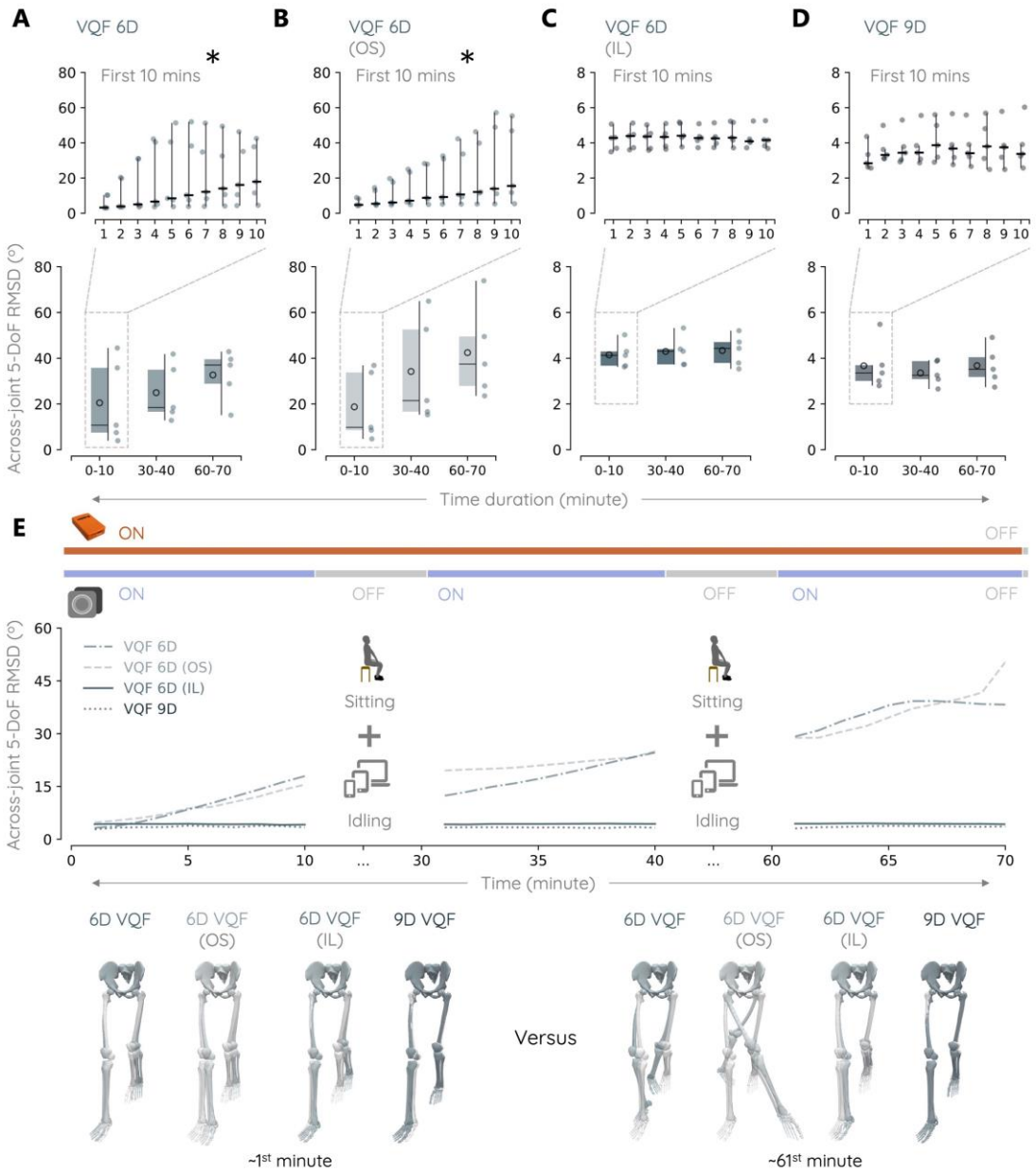

Figure 6. Impact of Biomechanical Constraints on Long-Term Drift (H2.B2 – Experiment 3). A) Due to drift, the 6D VQF showed a high median RMSD of above 10° even in the first 10 minutes of walking and reached nearly 40° in the last 10 minutes. **B)** Kinematics obtained from OpenSense (OS) also suffered from significant drift, with the median RMSD reaching nearly 50° during the last 10 minutes of walking. **C)** Our proposed approach IMoveLab (IL), in contrast, was able to achieve drift-free kinematics across the three walking trials taking place over 70 minutes. **D)** When the magnetometer information was available, the 9D VQF also maintained median RMSDs of less than 4° compared to the marker-based reference across the entire 70-minute experiment. **E)** All the IMU-based methods achieved kinematics that had a median agreement of 5° or better with marker-based kinematics at the beginning. However, only IMoveLab

ARTICLE IN PRESS

applied to 6D filters and the 9D filters were able to maintain the same agreement until the end of the 70-minute walking trial. The RMSDs in (A)–(D) were computed for the entire 10 minutes of walking at each trial. Each data point represents a participant (n = 5). Friedman tests were used to evaluate the overall effect of the time on RMSD, followed by two-sided Wilcoxon signed-rank test for post-hoc pairwise comparisons. The large vertical bars in (A)–(D) indicate the min-max range of the box plot, or Q1 – 1.5 * Interquartile range (IQR) to Q3 + 1.5 * IQR. Circles within the box plots denote the mean, while horizontal bars denote the median. Asterisks (*) indicate statistically significant differences, after applying Holm-Bonferroni correction (Friedman $p < 0.025$). Kinematics from only five degrees of freedom were considered in this analysis, with the non-sagittal-plane kinematics of the knees and ankles excluded.

Run Time Assessment

Next, we analyzed the computational time needed to obtain IMU-based kinematics of the lower body using different approaches: baseline filtering with accelerometer and gyroscope data (6D), biomechanically constrained 6D filtering (IMoveLab and OpenSense), and 9D filtering that contains magnetometer data for heading drift correction. This analysis was conducted on a workstation with an AMD Ryzen 9 5900X CPU (12 cores, 24 threads). Baseline 9D filters and 6D IMoveLab showed a run time of less than 3 ms per sample, but the latter took double the time (Fig. 7A). Constraint enforcement in IMoveLab was not only more accurate, as shown in earlier analyses, but also significantly faster than OpenSense, whose computational requirements increased polynomially with trial duration (Fig. 7B). For example, for a recording of about 2.5 minutes, IMoveLab took less than 10 seconds in total, while OpenSense required approximately 70 seconds. For hour-long trials (Experiment 3), IMoveLab generally required less than 20 minutes, whereas OpenSense required approximately 10 hours.

ARTICLE IN PRESS

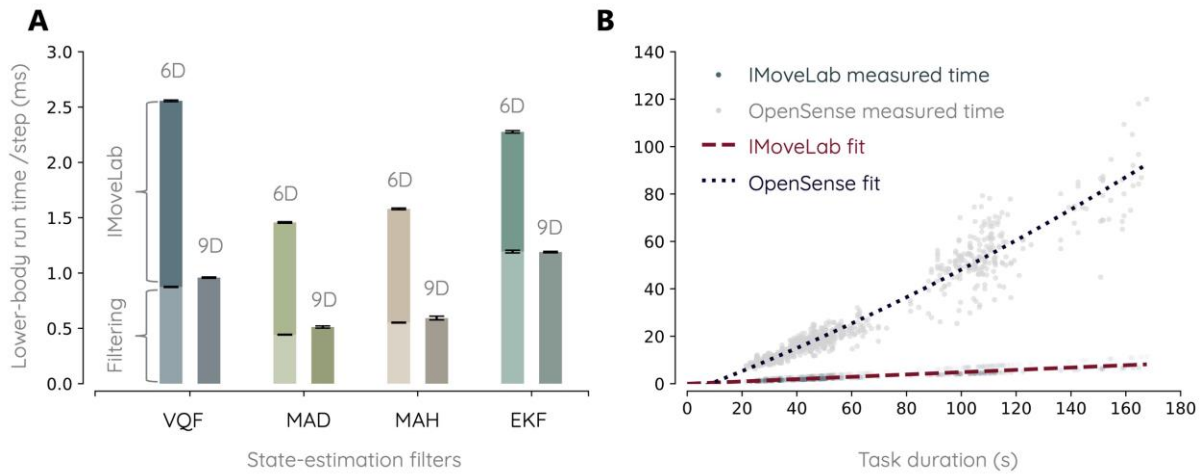

Figure 7. Measured Run Time by Approach (Auxiliary Analysis – Experiment 2). A) IMoveLab estimates leg kinematics in less than 3 ms per timestep, across different filters, indicating that IMoveLab can be used for real-time inference. The bars here indicate mean ± standard deviations from 19 participants. Individual data points are omitted given the small standard deviations relative to the size of the stacked bars. **B)** The time of constraint enforcement in IMoveLab scaled linearly with task duration, whereas that of OpenSense scaled quadratically. Each gray data point represents the run time of an individual trial when using OpenSense (n = 836, 4 filters, 11 tasks, 19 participants), while the dotted black line is a second-order polynomial fit to the data. The light blue data point represents the same run time when using IMoveLab, while the dashed red line represents a fit to these data.

Auxiliary Analysis: Effect of Covariates on Inertial- and Marker-Based-Tracking Agreement

We further investigated whether participant demographics, activity type, joint degree of freedom, and sensor placement influenced agreement between inertial and marker-based motion capture, using data from Experiment 2. We only conducted this auxiliary analysis on the 6D VQF constrained with IMoveLab and direct 9D VQF, as these approaches proved to be accurate and robust to drift. Sex, age, and body mass index (BMI) did not significantly impact IMoveLab's agreement with marker-based motion capture ($p > 0.026$, no significance after Holm-Bonferroni correction), but joint degree of freedom ($p < 0.001$), sensor placement ($p = 0.004$), and activity type ($p < 0.001$) did (Fig. 8). Hip flexion/extension had the highest agreement between the two motion capture approaches (median RMSD = 2.5° [2.1°, 2.7°]) compared to other degrees of freedom (hip adduction/abduction RMSD = 3.0° [2.3°, 3.6°], hip internal/external rotation RMSD = 4.6° [4.0°, 6.0°], knee flexion/extension RMSD = 3.4° [2.7°, 5.1°], ankle flexion RMSD = 3.3° [2.8°,

ARTICLE IN PRESS

4.1°], $p < 0.014$ for all comparisons). Sensors placed proximally on the shank led to estimates of knee flexion/extension that least approached results from the marker-based analysis compared to the distal-shank configurations ($p < 0.011$ for comparisons), but any median differences across configurations were less than 1° and likely practically insignificant when considering the accuracy of the marker-based system. IMoveLab exhibited higher disagreement with marker-based analysis in overground walking (median RMSD = 4.5° [3.9°, 5.2°]) compared to treadmill walking, treadmill running, drop jump, counter-movement jump, and step and hold ($p < 0.001$ for all comparisons) and in squatting (median RMSD = 4.4° [3.3°, 4.8°]) compared to treadmill walking, drop jump, counter-movement jump, and step and hold ($p < 0.001$ for all comparisons), with the median differences ranging from 0.8° to 1.5°. When using unconstrained 9D VQF, in addition to joint degree of freedom ($p < 0.001$), sensor placement ($p < 0.001$), and activity type ($p = 0.001$), sex ($p = 0.005$) also affected agreement between inertial and marker-based tracking, with disagreement being higher in females (Supplementary Fig. 4).

ARTICLE IN PRESS

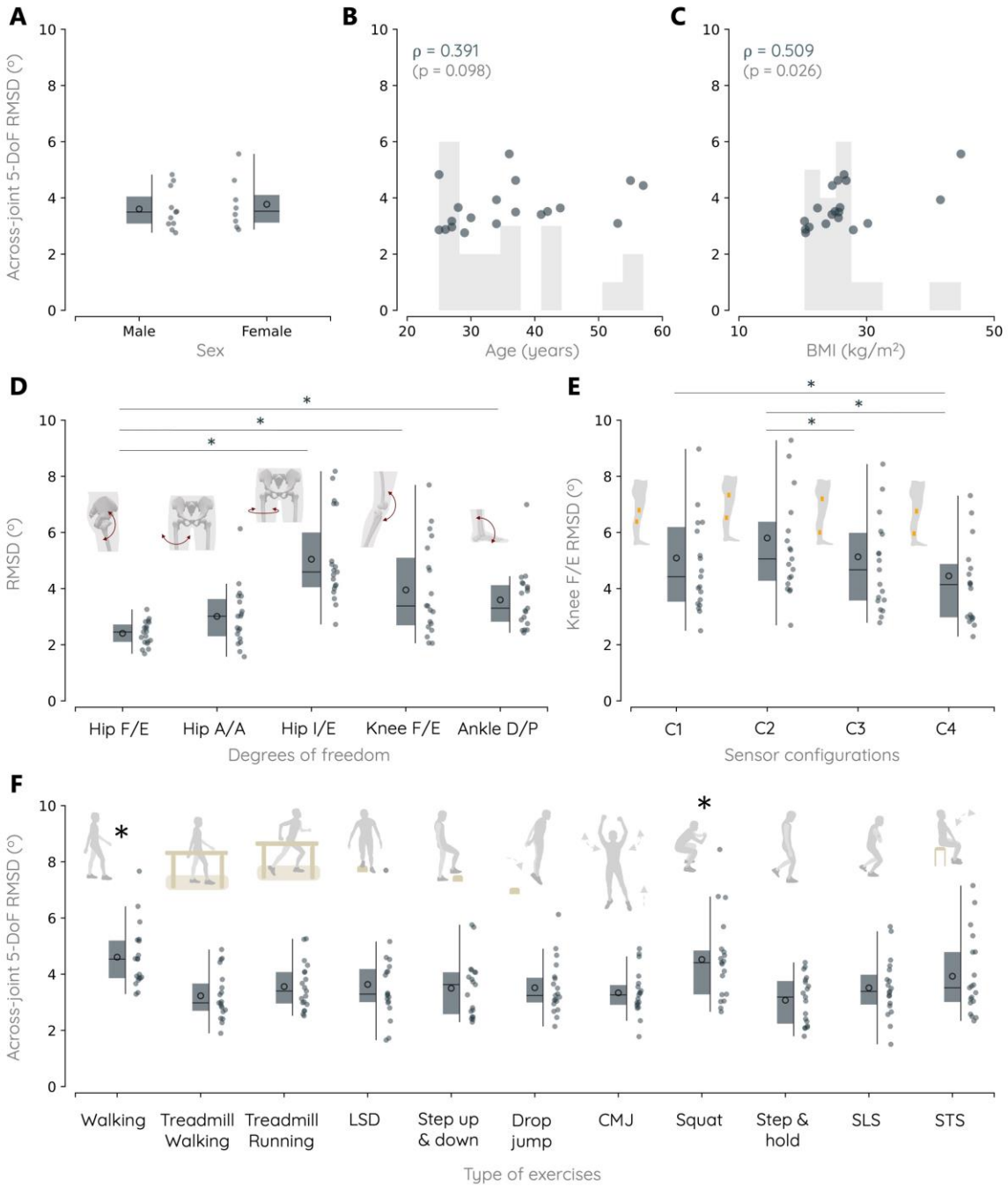

Figure 8. Impact of Covariates on 6D Constrained Kinematics with IMoveLab (Auxiliary Analysis – Experiment 2). Each data point represents a participant (n = 19). **A–C)** IMoveLab agreement with marker-based analysis did not differ by sex (Mann-Whitney U test), age, or BMI (Spearman correlation tests). **D)** RMSDs differed across degrees of freedom, with hip flexion/extension showing the strongest agreement and hip internal/external rotation showing the weakest agreement with marker-based analysis. **E)** Sensor placement affected RMSDs significantly, but by less than a degree. **F)** Across tasks, overground walking resulted in higher RMSDs compared to treadmill walking, treadmill

ARTICLE IN PRESS

running, drop jump, counter-movement jump, and step and hold; squatting also resulted in higher RMSD compared to treadmill walking, drop jump, counter-movement jump, and step and hold. Friedman tests were used to evaluate the overall effects of degree of freedom, sensor placement, and activity, followed by two-sided Wilcoxon signed-rank tests for post-hoc pairwise comparisons. Five degrees of freedom (5-DoF) in the lower body, including hip flexion/extension, hip adduction/abduction, hip internal/external rotation, knee flexion/extension, and ankle dorsi/plantarflexion, were considered in this analysis. F/E refers to flexion/extension, A/A abduction/adduction, I/E internal/external rotation, D/P dorsi/plantar, TM treadmill, CMJ counter-movement jump, SLS single-leg squat, and STS sit to stand. The large vertical bars in (A), (D), (E), and (F) indicate the min-max range of the box plot, or Q1 – 1.5 * Interquartile range (IQR) to Q3 + 1.5 * IQR. Circles within the box plots denote mean, while horizontal bars denote median. Asterisks (*) indicate statistically significant differences, after applying Holm-Bonferroni correction (degree-of-freedom Wilcoxon $p < 0.013$, sensor-placement Wilcoxon $p < 0.017$, and activity Wilcoxon $p < 0.001$).

DISCUSSION

Using biplane radiography as a ground-truth reference for bone kinematics, this study provides encouraging evidence that inertial tracking of the knee during walking and three jumping and hopping activities matches the accuracy of marker-based tracking, which for decades has been considered the de facto standard for evaluating human movement and, more recently, emerging wearable sensing technologies. Additionally, we proposed an approach to integrate biomechanical priors into the state-estimation loop to further improve the accuracy and eliminate the drift of IMU-based kinematics. Existing approaches, such as OpenSense, which use biomechanically constrained inverse kinematics, did not produce better estimates compared to baseline state-estimation filters in terms of accuracy and long-term drift. Our proposed approach, IMoveLab, maintained drift-free kinematics with 6D magnetometer-free filters, which other approaches can only achieve by adding magnetometer information.

This study is not without limitations. First, the findings reported here are encouraging, but they should be interpreted within the context of our experimental constraints, including the focus on walking, running, and orthopaedic-rehabilitation exercises. Future work may expand validation beyond the 14 activities studied here. Second, we focused on relatively healthy adults. Slight variations in accuracy are expected across populations, especially in patients who may not perform functional tasks for sensor-to-body calibration as

ARTICLE IN PRESS

well as those without mobility limitations. Third, we did not test 9D filters against biplane radiography due to severe magnetic disturbances inside the biplane-radiography system. In a marker-based laboratory environment, with treadmills and other ferromagnetic equipment typical of a gait lab, but without the biplane system, 9D VQF was robust to drift (Fig. 6). Our Experiments 2 and 3 took place in a gait lab with a non-negligible number of metal and electrical devices, and despite the lack of drift over time, we did observe that magnetic disturbances increased kinematics-estimation errors, especially for the ankle (Supplementary Fig. 5). On average, especially in outdoor environments, we do not expect daily use of 9D IMUs to be affected by magnetic disturbance more than it was in the laboratory. However, even if it were, IMoveLab could be combined with 9D filters to achieve drift-free kinematics. Fourth, although we tested the effects of sex, age, BMI, movement speed, joint degree of freedom, and sensor placement, we did not test for their interaction after considering statistical power. Fifth, in Experiment 2, we sampled inertial data at a lower frequency (40 Hz) than in Experiments 1 and 3 (100 Hz) since this experiment required the use of multiple IMUs per body segment for the sensor-placement sensitivity analysis. Xsens' restriction on wireless transmission did not allow higher sampling rates with multiple sensors per segment. Nevertheless, because the analyses related to hypothesis H2.B1 were performed using paired comparisons across methods within the same dataset, we do not expect the lower sampling frequency to have affected the conclusions regarding short-term drift.

ARTICLE IN PRESS

Another limitation of the study is that even though biplane radiography was used to obtain ground-truth skeletal kinematics in the first experiment, we only captured the kinematics of the knee. Biplane radiography has ionizing radiation, with small fields of view, short recording times, and time-consuming data processing requiring specialized expertise, which have historically limited the scaling of the technology across more biomechanics laboratories, despite its high accuracy and precision[32,33]. This may be one of the reasons why it has taken so long to perform an experimental validation like the one reported here, despite wide consensus on its necessity. While we have no reason to believe that marker-based analysis would be more accurate than inertial tracking in estimating ankle and hip kinematics, especially given the difficulty of marker placement around the hip, expanding the current validation to other joints remains a goal of future work. We believe that none of these limitations threatens the main conclusions reported here; instead, they

ARTICLE IN PRESS

provide a roadmap for the series of studies that could build on ours to take inertial tracking beyond the requisite scrutiny reserved for any emerging technology.

The long-duration experiment showcased that, in the absence of magnetometer information to mitigate heading drift, our proposed approach of strategically injecting biomechanical constraints into the state-estimation loop resolves a long-standing challenge in inertial tracking for human motion analysis. Since magnetometers may not always be available due to limitations in manufacturing or power consumption, short-range drift correction and machine learning have been explored as alternatives in the past. Rather than computing continuous kinematics, several prior studies have focused on range of motion per gait cycle as a primary outcome[7,34], achieving short-range drift correction by applying a step-detection algorithm and integrating only within a single stride. However, this approach is limited to cyclic activities, requires foot sensors for step detection, and has not been extensively validated, with only one prior study reporting sub-degree agreement with marker-based motion capture[34]. Machine learning approaches[15–17] have been viewed as a promising alternative, but they still suffer from lack of generalization out of sample, lack of published code, or the use of marker-based tracking as ground truth. It is therefore important to evaluate the literature on machine-learning solutions with appropriate caution until the code or data are shared. Unlike studies that propose reproducible algorithms that are data-agnostic, machine-learning models are heavily dependent on the data on which they were trained. Reproducing or verifying the validity of machine-learning solutions in the absence of the critical elements upon which their success hinges is therefore challenging. IMoveLab overcomes the limitations of both of these previous approaches.

Embedding biomechanical priors directly into the state estimation loop, rather than applying biomechanically constrained inverse kinematics sequentially after a state estimation filter, is advantageous because it alters how error evolves over time instead of simply constraining error magnitudes when they surpass the predefined physiological extremes. OpenSense treats the outputs of state-estimation filters as marker-based tracking, but these filters provide estimates. As a result, errors that accumulate are largely preserved when projected onto a constrained biomechanical model. In contrast, the step-by-step feedback formulation in IMoveLab prevents severe error accumulation over time. We do not use joint-limit constraints

ARTICLE IN PRESS

post hoc, which can allow significant error. We use informative biomechanical priors, such as coupling of degrees of freedom in the knee and spring mechanisms in the hip and ankle, to regulate the state trajectory within physiologically plausible manifolds, thereby preventing error accumulation. In addition to achieving high accuracy in lower-body motion tracking, our approach provides an advantage in computational time, as it only involves quaternion operations with constant time complexity. The observed differences in computational efficiency between OpenSense and IMoveLab are due to the polynomial time complexity of the least-squares optimization in OpenSense.

The finding that knee inertial tracking has comparable accuracy to marker-based tracking carries implications not only for technology adoption, but also for continued innovation and validation of new computational data-processing approaches. We showed that marker-based analysis can continue to serve as a reasonable reference standard for inertial-tracking approaches with lower accuracy than marker-based analysis, but not for methods with comparable accuracy (Supplementary Fig. 2). To date, developments in inertial tracking have aimed to reproduce marker-based motion tracking, with some prior inverse-kinematics or machine-learning tools reporting agreement below the technology's error margins[15,16]. While it is now evident that overfitting to a suboptimal "ground truth" cannot guide the field forward, these prior studies have offered a blueprint toward more imaginative integration of state-estimation filters, deep learning, and biomechanical modeling. Here we proposed one such approach, but continued innovation in fusion of complementary methods may continue to grow. This integration could lead inertial tracking to track skeletal kinematics even more accurately than reported here, across joints, especially when combined with historical biplane radiography data that offer a good prior for skin motion artifact. Until then, it is compelling enough to recognize that inertial tracking may not be as inferior to marker-based tracking as previously thought. Although additional experiments that extend validation across other joints will complement the findings reported here, this work provides a bracing glimpse into the future we have been cautious to leap into. We hope this glimpse fuels further experimentation and methodological innovation that continue to solidify the transition to widespread, confident use of inertial tracking.

ARTICLE IN PRESS

ARTICLE IN PRESS

METHODS

Human Participants and Ethics

In Experiment 1, we recruited 19 healthy athletes after obtaining approval from the Institutional Review Board at University of Pittsburgh (STUDY19080016). Only 15 participants (8 males and 7 females; age: 20.07 ± 1.24 years; height: 173.07 ± 9.99 cm; weight: 71.88 ± 10.86 kg) were included in the final analysis due to technical issues with four of them, leading to lack of data for at least one of the three motion-tracking modalities: inertial, marker-based, or biplane radiography. Two of the 15 participants were held out for tuning the state-estimation filters, while the remaining 13 were used for testing. In Experiment 2, we recruited 23 healthy adults after obtaining approval from the Institutional Review Board at Carnegie Mellon University (STUDY2022_00000178). Exclusion criteria were pregnancy and severe mobility limitations that limit independence (e.g., paralysis). Only 21 participants (9 females and 12 males; age: 35.10 ± 10.26 years; height: 172.67 ± 9.30 cm; weight: 77.37 ± 19.47 kg) were included in the final analysis due to technical issues with two of them, leading to unreliable IMU data. Two of the 21 participants were held out for tuning the state-estimation filters, while the remaining 19 were used for testing. Five of these participants also completed a long-duration trial (Experiment 3). Each participant in these studies provided written informed consent to participate in the study but not to disclose personally identifiable information. Participants were compensated at an hourly rate.

ARTICLE IN PRESS

Experiment 1: Accuracy Evaluation against Biplane Radiography

To compare inertial- and marker-based kinematics against ground-truth skeletal kinematics, we used a biplane-radiography system[35] that captured knee kinematics while participants performed four tasks: single-leg hop with 180° internal rotation, single-leg drop, double-leg drop jump, and treadmill running (Figs. 1A and B). Each task was recorded three times for each limb, summing up to six trials per task. Inertial data (BioStamp nPoint, MC10 Inc., Lexington, MA, USA) were captured during the entire experiment without interruption. Marker-based data (Vicon Vantage V5, Vicon Motion Systems Inc., Oxford, UK) were captured for 10 seconds per trial. Three seconds after the start of the marker-based system, the biplane system[36] was triggered to capture radiographs of the knee. Due to the presence of ionizing radiation, biplane-radiography data were only captured for a fraction of a second (up to one second). A modified Rizzoli

ARTICLE IN PRESS

markerset, which included segment clusters to reduce soft-tissue artifact and improve tracking accuracy, was used to track the lower body with 12 cameras, at a sampling rate of 100–150 Hz. Four IMUs were directly attached to the thighs and shanks using BioStamp double-sided adhesives. These IMUs captured data at 250 Hz. The marker-based tracking clusters were placed on top of the IMUs using self-adherent elastic wrap (3M Coban, 3M, St. Paul, MN, USA). Biplane radiographs were captured at 150 Hz, with 1 ms exposure. Biplane and marker-based data were hardware synchronized. Inertial data and marker-based data were synchronized by aligning the Unix timestamps of the two systems and then correcting with the cross-correlation of the knee flexion angles. Data from this experiment were used to test hypothesis 1 (H1) that marker-based analysis is more accurate than inertial tracking and hypothesis 2.A (H2.A) on the effect of biomechanical constraints on inertial-tracking accuracy. When comparing accuracies, only identical segments of time-synchronized data across the three technologies were used.

Experiment 2: Drift Evaluation in Short Trials across Eleven Tasks

Participants were instructed to perform three locomotion tasks—including overground walking, treadmill walking, and treadmill running—and eight common lower-extremity exercises (Fig. 2, and Supplementary Table 1). These movements were selected to span a range of common dynamic activities encountered in daily life and orthopaedic-rehabilitation assessments. These tasks also represent a range of movement speeds and joint excursions to test robustness under different conditions relevant to orthopaedic rehabilitation. Locomotion activities were recorded for one to two minutes each, whereas the functional exercises were repeated three times, with unilateral exercises (e.g., lateral step-down or step-and-hold) performed with each limb. Motion data were collected using both inertial (MTw Awinda, Movella, Henderson, NV, USA) and optical (OptiTrack NaturalPoint, Corvallis, OR, USA) motion-tracking systems (Fig. 2A), with the latter used to obtain reference kinematics. A modified Rizzoli markerset was used to track the lower body with 20 cameras and a sampling frequency of 100 Hz. Fifteen IMUs placed on the pelvis, thighs, shanks, and feet captured data at 40 Hz using a free-license software (MT Manager, Movella, Henderson, NV, USA). Redundant sensors were placed on the thighs and shanks to assess the impact of sensor placement on kinematics. We asked participants to hop three times before each task for synchronization of the IMU-based and marker-based systems. Data from this experiment were used to test

ARTICLE IN PRESS

hypothesis 2.B1 (H2.B1), on the role of biomechanical constraints on short-term inertial-tracking drift across more tasks and joints than allowed by Experiment 1. They were also used for auxiliary analyses testing the effects of sex, age, BMI, degree of freedom, sensor placement, and activity type on agreement between inertial- and-marker-based tracking.

Experiment 3: Drift Evaluation in a Long-Duration Walking Trial

To test hypothesis 2.B2 (H2.B2) on the effect of biomechanical constraints on longer-term drift, we performed a long-duration experiment, wherein a subset of five volunteers from Experiment 2 completed a 70-minute experiment (Fig. 2C). Seven IMUs were used to capture data continuously at 100 Hz. Participants performed three sequences of 10-minute walking bouts followed by 5 minutes of sitting, with each sequence separated by a 15-minute break, during which participants engaged in activities of their choice (e.g., phone or laptop use). Walking was selected given that it is one of the most common and functionally relevant daily activities, widely studied in biomechanics[37] and assistive-device research[38–40]. Marker-based motion capture was turned off during the breaks. We asked participants to hop three times before the task for synchronization of the IMU-based and marker-based systems, repeating the three hops each time the marker-based system was reactivated during the long-duration trial.

ARTICLE IN PRESS

Data Processing and Analysis

In Experiment 1, all the data collected from biplane-radiography, inertial, and marker-based systems were resampled to 100 Hz before processing. In Experiments 2 and 3, all the data (i.e., marker trajectories and inertial data) were resampled to 40 Hz and 100 Hz, respectively, to match the IMU sampling rate. The lower sampling rate in Experiment 2 was due to Xsens hardware (i.e., MTw Awinda) limitations on wireless transmission, given that we used multiple sensors per limb to assess the effect of sensor placement.

To process the biplane radiography data (Experiment 1), participant-specific three-dimensional models of the femur and tibia were obtained from computed tomography (CT) scans with an average in-plane resolution of 0.37 x 0.3 mm and slice thickness of 0.625 mm (LightSpeed Pro 16; GE Medical Systems, Waukesha, WI), segmented semi-automatically (Mimics software, Materialise Inc, Ann Arbor, MI). The

ARTICLE IN PRESS

average radiation exposure from the CT scan was 1.27 mSv, and the exposure from the biplane radiography system was estimated to be 0.27 mSv. The anatomical coordinate systems of these bone models were determined using an automated algorithm based on articular surface geometry[41] and single CT slices through the proximal femur and distal tibia[42]. Ground-truth knee kinematics were obtained by matching the bone models with the biplane radiographs through an iterative optimization process, wherein digitally reconstructed radiographs created from subject-specific CT bone models best matched each pair of synchronized radiographs using a validated technique[33]. Biplane radiography has been previously shown to have sub-millimeter and sub-degree accuracy in cadaver studies[32]. Tracking precision for dynamic knee rotations in human participants has been reported to be 0.9° ± 0.3°, 0.6° ± 0.3°, and 0.3° ± 0.1° for flexion/extension, external/internal rotation, and ab/adduction, respectively[33]. Once obtained, kinematics were filtered using a fourth-order Butterworth low-pass filter, with a cut-off frequency of 8.5 Hz.

To obtain kinematics from the marker-based systems (Experiments 1–3), marker trajectories were filtered using a fourth-order Butterworth low-pass filter, with a cutoff frequency of 15 Hz. During static standing, joint coordinate systems for the marker-based analysis were identified using anatomical marker data, following the Grood and Suntay convention[12,43,44]. Transformations from tracking to anatomical coordinate systems were used to obtain kinematics from the tracking clusters, using an unconstrained kinematics approach.

We selected five state-estimation filters (Table 1) for estimation of lower-body-segment orientations from the IMUs. The MAD[18] and MAH[19] filters were selected because they are the most popular specialized complementary filters, the EKF[23] because it is the most widely used Bayesian filter, the VQF[10,24] because it is regarded as the most accurate adaptive filter, and the machine-learning filter RIANN[17] for its generalizability without the need for fine-tuning. Due to strong magnetic disturbances caused by the electrical components of the biplane radiography system, we used only gyroscope and accelerometer data (6D) for kinematics estimation from these filters. Data from standing, walking, and jumping trials were used for sensor-to-segment calibration, following previously presented approaches[34,45] (Supplementary Fig. 6). More specifically, by assuming that the proximal-distal axes of the body segments are aligned with gravity

ARTICLE IN PRESS

during static standing, we identified the proximal-distal axis (Y) of each segment, expressed in the sensor frame, by averaging the three-axis accelerometer data of the static pose. We then identified a temporary medio-lateral axis (Z') by applying principal component analysis (PCA) to the three-axis gyroscope data of eight steady-state gait cycles taken from straight walking on the treadmill. The anterior-posterior axis (X) of each segment was then obtained via the cross product of the proximal-distal (Y) and temporary medio-lateral (Z') axes. To ensure rotation-matrix orthogonality, the medio-lateral (Z) axis was computed via the cross product of the proximal-distal (Y) and anterior-posterior (X) axes.

Table 1. List of State-Estimation Filters.

State-estimation filter	Acronym	Gyr.	Acc.	Mag.
Madgwick filter	MAD	✔	✔	Optional
Mahony filter	MAH	✔	✔	Optional
Extended Kalman filter	EKF	✔	✔	Optional
Versatile quaternion-based filter	VQF	✔	✔	Optional
Robust IMU-based attitude neural network	RIANN	✔	✔	–

Unconstrained kinematics were computed directly using the segment orientations derived from the state-estimation filters. For constrained kinematics, we used IMoveLab, which applies biomechanics constraints recursively at each timestep and provides continuous feedback to a state-estimation filter (i.e., sensor-fusion algorithm), along with OpenSense[25], which applies inverse kinematics with the Rajagopal model[46] sequentially after the application of filters. We sought to eliminate the effect of different joint coordinate systems across the different motion-tracking technologies by identifying the transformations across segment-specific anatomical coordinate systems in static trials and applying these transformations to dynamic trials. For Experiment 1, the anatomical-coordinate-system differences across the three technologies were identified using data from the beginning of the biplane-radiography capture. For Experiments 2–3, differences between the inertial and marker-based coordinate systems were identified during the static pose. Unconstrained and IMoveLab kinematics were obtained using quaternion operations. Euler angles were only used in the final reporting of joint kinematics[12,14].

For all experiments, we computed root-mean-square difference (RMSD) against a reference (e.g., biplane radiography in Experiment 1 or marker-based analysis in Experiments 2 and 3). RMSDs between inertial

ARTICLE IN PRESS

tracking and biplane radiography, marker-based analysis and biplane radiography, and inertial tracking and marker-based analysis were computed over the same specific movements. For Experiment 1, data from inertial and marker-based analyses were trimmed to match the length of biplane-radiography kinematics. For the short-duration drift analysis in Experiment 2, we used change in RMSD as the primary outcome, based on the definition of drift. RMSDs from the first and last five gait cycles were used for locomotion tasks, while those from the first and last repetition were used for other exercises.

IMoveLab: Coupling Constraints (Knee)

Considering the Markovian property of state-estimation filters (i.e., estimates in the next timestep depend on the current one), we implemented an approach that integrates knee coupling information recursively into the filter's state-estimation loop (Fig. 1E, Supplementary Method 1), instead of only enforcing joint-limit constraints sequentially after at the end of the state-estimation process without intermediate feedback, as is the case in OpenSense. Specifically, secondary degrees of freedom in the knee can be parameterized as polynomial functions of knee flexion/extension[47], which is estimated more accurately than ab/adduction and internal/external rotation. The error between the estimated and constraint-informed kinematics of these secondary degrees of freedom is then propagated back to correct the orientation of the distal sensor (i.e., shank sensor) before the filter updates in the next timestep. This correction is enforced via a soft constraint with a tunable gain α, where $0 \leq \alpha \leq 1$, allowing the strength of the coupling prior to be adjusted for different applications and populations. Setting α closer to 1 will enforce the coupling constraints more strictly, while setting α closer to 0 will favor the direct (or unconstrained) kinematics. We empirically set $\alpha = 0.9$ across all the experiments. While the constraints focus on the secondary degrees of freedom, this approach inherently corrects inaccuracies and drift in knee flexion/extension as well. Importantly, the nominal coupling relationship used here is modular and can be enhanced with population-specific data (e.g., people after anterior-cruciate-ligament reconstruction).

IMoveLab: Spring-Like Muscle–Tendon Constraints (Hip and Ankle)

Because explicit kinematic coupling is less well-defined at the hip and ankle joints, we instead drew from the spring-like behavior of muscle–tendon units to introduce a weak corrective term that mitigates heading

ARTICLE IN PRESS

drift (Fig. 2D, Supplementary Method 2), especially when kinematics estimation is performed without heading information (e.g., in the absence of magnetometers). Specifically, the error term for correction of joint internal/external rotation is defined as a small proportion of the current rotation estimate, introducing a weak restorative effect that continuously biases the joint toward a neutral pose, defined here as 0°. A parameter, $0 \leq \kappa \leq 1$, was introduced to control the strength of this enforcement. Higher values of κ (close to 1) correspond to stiffer springs, constraining the joint near the predefined neutral pose, which is appropriate when internal/external rotation is not of primary interest. Lower values of κ (close to 0) correspond to more compliant springs, allowing physiological motion while still attenuating drift, which is preferable when secondary degrees of freedom are of interest. We empirically selected $\kappa = 0.0064$ for the hip and $\kappa = 0.25$ for the ankle. While this scheme performed robustly across a range of movements, it conservatively biases joint rotation to be 0° during static poses. This behavior is intentional to avoid the drift amplification during static poses in the absence of heading information.

Statistical Analysis

To test the first hypothesis (H1) that marker-based analysis is more accurate than inertial tracking, we applied multiple Wilcoxon signed-rank tests for pairwise comparisons across methods, degrees of freedom, and tasks. We also conducted multiple Spearman correlation tests to assess the association between inertial tracking agreement with biplane radiography (error) and agreement with marker-based analysis. To test the second hypothesis (H2.A) that our proposed approach, IMoveLab, achieves higher accuracy than both direct state estimation with baseline filters and OpenSense, we used Friedman tests, followed by Wilcoxon signed-rank tests for multiple comparisons. To test H2.B1 that IMoveLab is better able to reduce short-duration drift than both baseline filters and OpenSense, we also used Friedman tests, followed by Wilcoxon signed-rank tests for multiple comparisons. To test H2.B2 that IMoveLab is better able to reduce long-duration drift than both baseline filters and OpenSense, we also used Friedman tests, followed by Wilcoxon signed-rank tests for multiple comparisons. We investigated the effects of sex on inertial- and marker-based-tracking agreement by applying a Mann-Whitney U test; the effects of age and BMI via Spearman correlation tests; and the effects of degree of freedom and sensor placement using Friedman tests, followed by Wilcoxon signed-rank tests for possible pairwise comparisons. To account for multiple

ARTICLE IN PRESS

comparisons, we used Holm-Bonferroni correction, where applicable. The results are reported as median [Q1, Q3]. Non-parametric tests were selected because the assumptions required for parametric testing were not met by most outcome variables.

Data Availability Statement: All data supporting the findings of this study are available within the article and its supplementary files. Any additional requests for information can be directed to, and will be fulfilled by, the corresponding authors. Source data are provided with this paper. The raw data generated in this study have been deposited in Figshare: https://figshare.com/s/95a87e71c2e3f252a07f?file=60333584.

Code Availability Statement: The code is available on GitHub: https://github.com/CMU-MBL/IMoveLab. An IMoveLab graphical user interface is also being actively developed as a public resource, with a preliminary prototype found here: https://imovelab.org/.

ARTICLE IN PRESS

ARTICLE IN PRESS

REFERENCES

1. Friesen, K. B., Zhang, Z., Monaghan, P. G., Oliver, G. D. & Roper, J. A. All eyes on you: how researcher presence changes the way you walk. *Sci. Rep.* **10**, 17159 (2020).
2. Phan, V., Meinders, E., Kiani, R., Sprague, A. L., Irrgang, J. J., Hughes, J. D., Rabuck, S. J., Lesniak, B. P., Musahl, V. & Halilaj, E. The gait lab effect: symmetry restoration strategy after anterior cruciate ligament reconstruction is different in natural environments than the gait laboratory. *J. Biomech.* 113379 (2026) doi:10.1016/j.jbiomech.2026.113379.
3. Parola, L. R., Phan, V. & Halilaj, E. Laboratory Environments Mask Gait Differences between Healthy Participants and Patients with Anterior Cruciate Ligament Reconstruction. *IEEE Trans. Neural Syst. Rehabil. Eng.* 1–1 (2025) doi:10.1109/TNSRE.2025.3645799.
4. Wellsandt, E., Gardinier, E. S., Manal, K., Axe, M. J., Buchanan, T. S. & Snyder-Mackler, L. Decreased knee joint loading associated with early knee osteoarthritis after anterior cruciate ligament injury. *Am. J. Sports Med.* **44**, 143–151 (2016).
5. Erhart-Hledik, J., Chu, C., Asay, J., Favre, J. & Andriacchi, T. Longitudinal Changes in the Total Knee Joint Moment After Anterior Cruciate Ligament Reconstruction Correlate With Cartilage Thickness Changes. *J. Orthop. Res. Off. Publ. Orthop. Res. Soc.* **37**, 1546–1554 (2019).
6. Ray, T., Choi, J., Reeder, J., Lee, S. P., Aranyosi, A. J., Ghaffari, R. & Rogers, J. A. Soft, skin-interfaced wearable systems for sports science and analytics. *Curr. Opin. Biomed. Eng.* **9**, 47–56 (2019).
7. Kobsar, D., Masood, Z., Khan, H., Khalil, N., Kiwan, M. Y., Ridd, S. & Tobis, M. Wearable Inertial Sensors for Gait Analysis in Adults with Osteoarthritis—A Scoping Review. *Sensors* **20**, 7143 (2020).
8. Kobsar, D., Charlton, J. M., Tse, C. T. F., Esculier, J.-F., Graffos, A., Krowchuk, N. M., Thatcher, D. & Hunt, M. A. Validity and reliability of wearable inertial sensors in healthy adult walking: a systematic review and meta-analysis. *J. NeuroEngineering Rehabil.* **17**, 62 (2020).
9. Poitras, I., Dupuis, F., Bielmann, M., Campeau-Lecours, A., Mercier, C., Bouyer, L. J. & Roy, J.-S. Validity and Reliability of Wearable Sensors for Joint Angle Estimation: A Systematic Review. *Sensors* **19**, 1555 (2019).

ARTICLE IN PRESS

ARTICLE IN PRESS

10. Laidig, D. & Seel, T. VQF: Highly accurate IMU orientation estimation with bias estimation and magnetic disturbance rejection. *Inf. Fusion* **91**, 187–204 (2023).
11. Nazarahari, M. & Rouhani, H. Sensor fusion algorithms for orientation tracking via magnetic and inertial measurement units: An experimental comparison survey. *Inf. Fusion* **76**, 8–23 (2021).
12. Grood, E. S. & Suntay, W. J. A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. *J. Biomech. Eng.* **105**, 136–144 (1983).
13. Wu, G., Siegler, S., Allard, P., Kirtley, C., Leardini, A., Rosenbaum, D., Whittle, M., D'Lima, D. D., Cristofolini, L., Witte, H., Schmid, O. & Stokes, I. ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine. *J. Biomech.* **35**, 543–548 (2002).
14. Cereatti, A., Gurchiek, R., Mündermann, A., Fantozzi, S., Horak, F., Delp, S. & Aminian, K. ISB recommendations on the definition, estimation, and reporting of joint kinematics in human motion analysis applications using wearable inertial measurement technology. *J. Biomech.* **173**, 112225 (2024).
15. Mundt, M., Johnson, W. R., Potthast, W., Markert, B., Mian, A. & Alderson, J. A Comparison of Three Neural Network Approaches for Estimating Joint Angles and Moments from Inertial Measurement Units. *Sensors* **21**, 4535 (2021).
16. Rapp, E., Shin, S., Thomsen, W., Ferber, R. & Halilaj, E. Estimation of kinematics from inertial measurement units using a combined deep learning and optimization framework. *J. Biomech.* **116**, 110229 (2021).
17. Weber, D., Gühmann, C. & Seel, T. RIANN—A Robust Neural Network Outperforms Attitude Estimation Filters. *AI* **2**, 444–463 (2021).
18. Madgwick, S. O. H. An efficient orientation filter for inertial and inertial/magnetic sensor arrays. *Rep. X-Io Univ. Bristol UK* **25**, 113–118 (2010).
19. Mahony, R., Hamel, T. & Pflimlin, J.-M. Nonlinear Complementary Filters on the Special Orthogonal Group. *IEEE Trans. Autom. Control* **53**, 1203–1218 (2008).
20. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **82**, 35–45 (1960).

ARTICLE IN PRESS

ARTICLE IN PRESS

21. Kok, M., Hol, J. D. & Schön, T. B. Using Inertial Sensors for Position and Orientation Estimation. *Found. Trends® Signal Process.* **11**, 1–153 (2017).
22. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **82**, 35–45 (1960).
23. Sabatini, A. M. Kalman-filter-based orientation determination using inertial/magnetic sensors: observability analysis and performance evaluation. *Sensors* **11**, 9182–9206 (2011).
24. Laidig, D., Weygers, I., Bachhuber, S. & Seel, T. VQF: A Milestone in Accuracy and Versatility of 6D and 9D Inertial Orientation Estimation. in *2022 25th International Conference on Information Fusion (FUSION)* 1–6 (2022). doi:10.23919/FUSION49751.2022.9841356.
25. Al Borno, M., O'Day, J., Ibarra, V., Dunne, J., Seth, A., Habib, A., Ong, C., Hicks, J., Uhlrich, S. & Delp, S. OpenSense: An open-source toolbox for inertial-measurement-unit-based measurement of lower extremity kinematics over long durations. *J. NeuroEngineering Rehabil.* **19**, 22 (2022).
26. Roetenberg, D., Luinge, H. & Slycke, P. Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors. *Xsens Motion Technol. BV Tech. Rep.* **1**, 1–7 (2009).
27. Schepers, M., Giuberti, M. & Bellusci, G. Xsens MVN: Consistent Tracking of Human Motion Using Inertial Sensing. *Xsens Technol. BV* **1**, 1–8 (2018).
28. Thrun, S., Burgard, W. & Fox, D. *Probabilistic Robotics*. (MIT Press, Cambridge, MA, USA, 2005).
29. Miranda, D. L., Rainbow, M. J., Crisco, J. J. & Fleming, B. C. Kinematic differences between optical motion capture and biplanar videoradiography during a jump–cut maneuver. *J. Biomech.* **46**, 567–573 (2013).
30. Langworthy, M., Dasa, V. & Spitzer, A. I. Knee osteoarthritis: disease burden, available treatments, and emerging options. *Ther. Adv. Musculoskelet. Dis.* **16**, 1759720X241273009 (2024).
31. Singh, J. A., Yu, S., Chen, L. & Cleveland, J. D. Rates of Total Joint Replacement in the United States: Future Projections to 2020–2040 Using the National Inpatient Sample. *J. Rheumatol.* **46**, 1134–1140 (2019).
32. Miranda, D. L., Schwartz, J. B., Loomis, A. C., Brainerd, E. L., Fleming, B. C. & Crisco, J. J. Static and Dynamic Error of a Biplanar Videoradiography System Using Marker-Based and Markerless Tracking Techniques. *J. Biomech. Eng.* **133**, (2011).

ARTICLE IN PRESS

ARTICLE IN PRESS

33. Anderst, W., Zauel, R., Bishop, J., Demps, E. & Tashman, S. Validation of three-dimensional model-based tibio-femoral tracking during running. *Med. Eng. Phys.* **31**, 10–16 (2009).
34. Hafer, J. F., Provenzano, S. G., Kern, K. L., Agresta, C. E., Grant, J. A. & Zernicke, R. F. Measuring markers of aging and knee osteoarthritis gait using inertial measurement units. *J. Biomech.* **99**, 109567 (2020).
35. Nishida, K., Xu, C., Gale, T., Anderst, W. & Fu, F. Symmetry and sex differences in knee kinematics and ACL elongation in healthy collegiate athletes during high-impact activities revealed through dynamic biplane radiography. *J. Orthop. Res.* **40**, 239–251 (2022).
36. Tashman, S., Princehorn, J., Pennatto, S. & Anderst, W. Bi-plane X-ray imaging system. (2010).
37. Whittle, M. W. Clinical gait analysis: A review. *Hum. Mov. Sci.* **15**, 369–387 (1996).
38. Slade, P., Kochenderfer, M. J., Delp, S. L. & Collins, S. H. Personalizing exoskeleton assistance while walking in the real world. *Nature* **610**, 277–282 (2022).
39. Molinaro, D. D., Kang, I. & Young, A. J. Estimating human joint moments unifies exoskeleton control, reducing user effort. *Sci. Robot.* **9**, eadi8852 (2024).
40. Kim, J., Porciuncula, F., Yang, H. D., Wendel, N., Baker, T., Chin, A., Ellis, T. D. & Walsh, C. J. Soft robotic apparel to avert freezing of gait in Parkinson's disease. *Nat. Med.* **30**, 177–185 (2024).
41. Renault, J.-B., Aüllo-Rasser, G., Donnez, M., Parratte, S. & Chabrand, P. Articular-surface-based automatic anatomical coordinate systems for the knee bones. *J. Biomech.* **80**, 171–178 (2018).
42. Tashman, S. & Anderst, W. In-Vivo Measurement of Dynamic Joint Motion Using High Speed Biplane Radiography and CT: Application to Canine ACL Deficiency. *J. Biomech. Eng.* **125**, 238–245 (2003).
43. Wu, G. & Cavanagh, P. R. ISB recommendations for standardization in the reporting of kinematic data. *J. Biomech.* **28**, 1257–1261 (1995).
44. Wu, G., Siegler, S., Allard, P., Kirtley, C., Leardini, A., Rosenbaum, D., Whittle, M., D'Lima, D. D., Cristofolini, L., Witte, H., Schmid, O., Stokes, I., & Standardization and Terminology Committee of the International Society of Biomechanics. ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion--part I: ankle, hip, and spine. International Society of Biomechanics. *J. Biomech.* **35**, 543–548 (2002).

ARTICLE IN PRESS

45. Nazarahari, M. & Rouhani, H. Semi-Automatic Sensor-to-Body Calibration of Inertial Sensors on Lower Limb Using Gait Recording. *IEEE Sens. J.* **19**, 12465–12474 (2019).
46. Rajagopal, A., Dembia, C. L., DeMers, M. S., Delp, D. D., Hicks, J. L. & Delp, S. L. Full-Body Musculoskeletal Model for Muscle-Driven Simulation of Human Gait. *IEEE Trans. Biomed. Eng.* **63**, 2068–2079 (2016).
47. Walker, P. S., Rovick, J. S. & Robertson, D. D. The effects of knee brace hinge design and placement on joint mechanics. *J. Biomech.* **21**, 965–974 (1988).

Acknowledgements: This work was funded by the National Science Foundation CAREER Award (CBET 2145473) and the Dowd Fellowship from the College of Engineering at Carnegie Mellon University. We would like to thank the volunteers for their time participating in this study.

Funding Statement: We disclose support for the research of this work from the National Science Foundation [CBET 2145473] and the Dowd Fellowship from the College of Engineering at Carnegie Mellon University.

Author Contributions: Conceptualization: V.P., Z.L., and E.H.; Methodology: V.P., T.G., W.A., and E.H.; Experimentation: V.P., Z.L., E.M., T.G., and A.K.; Formal Analysis: V.P., A.K., T.G., and E.H.; Investigation: V.P., W.A., and E.H.; Visualization: V.P., T.G., and E.H.; Software: V.P. and J.N.; Writing – original draft: V.P. and E.H.; Writing – review and editing: all authors. Supervision: W.A. and E.H.; Funding Acquisition: E.H.

Competing Interests Statement: V.P., E.M., and E.H. are co-inventors on a provisional patent application related to the biomechanics-informed sensor-fusion framework described here, filed by Carnegie Mellon University. The remaining authors declare no competing interests.