Degree Project in Sport Technology

Second Cycle, 30 credits

# Development and Validation of a Power Meter for Functional Fitness Athletes

**NICOLÒ RUBELE**

*This page intentionally left blank.*

# Development and Validation of a Power Meter for Functional Fitness Athletes

Nicolò Rubele

Master's Programme, Sports Technology
Supervisor: Mariah Sabioni
Reviewer: Jonas Willén
Examiner: Sebastiaan Meijer
School of Engineering Sciences in Chemistry, Biotechnology and Health
Swedish title: Utveckling och Validering av en Effektmätare för Idrottare med Functional Fitness
TRITA - CBH-GRU 2025:041

*This page intentionally left blank.*

# Abstract

Power meters have been available in endurance sport since the 1980s and provide useful insights to individualized training programs in order to improve the effectiveness of training sessions. This also helps in reducing injury risks. In gym context, especially for functional fitness (FF), nowadays it is not possible to measure the power output (PO) of athletes, thus making it difficult to understand how different training sessions can affect athletes' bodies, both in terms of performance and fatigue.

The main aim of this study is to develop a new inertial measurement unit (IMU)-based power meter to be used in FF contexts and to validate it against a gold standard, a combination of a motion capture system (MoCap) and 4 force plates. Additionally, an investigation about the best configuration of sensors to be used is carried out.

Seven well trained FF athletes took part in this study. They were asked to perform some FF movements (burpees, clean and jerks, lunges, snatches and thrusters) while being recorded by 3 IMUs placed on them (chest, wrist and ankle) and by the gold standard system. A Python algorithm applied to IMU data estimated PO and this was compared with the PO computed from the gold standard. The new power meter estimated the PO with two different approaches: the force approach (FA), based on the formula P=F*v, and the energy approach (EA), based on mechanical energy.

Both FA and EA showed very high to almost perfect correlation with the gold standard, with the best correlation (0.927) provided by FA with the configuration with 2 sensors at chest and wrist, when comparing the full PO curve. Analyzing average positive and negative power, total positive and negative work and peak positive and negative power, FA showed better estimations despite a general underestimation of around 10% for all the metrics analyzed and with a RMSE around 25% (after the correction of the raw estimations, while raw PO overestimates the MoCap by around 33%).

The correlations between IMU and MoCap found in this study are in line with the values found by other studies. A comparison of the other metrics with literature is difficult due to a lack of similar types of studies. Despite that, studies comparing IMU and MoCap to estimate PO during countermovement jumps found a percentage overestimation around 25%, close to the raw estimation of around 33% of this study.

This study showed that the best method to estimate PO with IMUs is FA. Moreover, even if the 3-IMU power meter showed promising results, the configuration with just two sensors (at chest and wrist) provided slightly better estimations.

**Keywords**
IMU, MoCap, Functional Fitness, CrossFit, power meter, validation

*This page intentionally left blank.*

Sammanfattning

Effektmätare har funnits inom uthållighetsidrott sedan 1980-talet och ger värdefulla insikter för individanpassade träningsprogram i syfte att förbättra effektiviteten i träningspassen. Detta bidrar också till att minska risken för skador. Inom gymmiljöer, särskilt inom funktionell fitness (FF), är det idag inte möjligt att mäta idrottares effektutveckling (PO), vilket gör det svårt att förstå hur olika träningspass påverkar kroppen, både vad gäller prestation och trötthet.

Det huvudsakliga syftet med denna studie är att utveckla en ny effektmätare baserad på inertialmätsystemer (IMU) för användning inom FF och att validera den mot en guldstandard, en kombination av ett rörelseanalyssystem (MoCap) och 4 kraftplattor. Dessutom genomförs en undersökning av vilken konfiguration som är mest lämplig.

Sju vältränade FF-atleter deltog i studien. De ombads utföra olika FF-rörelser (burpees, clean and jerks, lunges, snatches och thrusters) samtidigt som de registrerades av tre IMU-enheter placerade på bröstkorg, handled och fotled samt av guldstandardsystemet. Ett Python-algoritm applicerades på IMU-datan för att uppskatta PO, vilken sedan jämfördes med PO beräknad från guldstandarden. Den nya effektmätaren uppskattade PO med två olika metoder: kraftmetoden (FA), baserad på formeln P = F*v, och energimetoden (EA), baserad på mekanisk energi.

FA och EA visade mycket hög till nästan perfekt korrelation med guldstandarden, med den bästa korrelationen (0,927) för FA med konfigurationen med två sensorer på bröst och handled vid jämförelse av hela effektkurvan. Vid analys av genomsnittlig positiv och negativ effekt, total positiv och negativ arbetsinsats samt toppvärden för positiv och negativ effekt visade FA bättre resultater, trots en generell underskattning på cirka 10% för alla analyserade mätvärden och med ett RMSE på cirka 25% (efter korrigering av den råa PO, överestimerar den råa PO med cirka 33% jämfört med MoCap).

Korrelationerna mellan IMU och MoCap som hittades i denna studie ligger i linje med de värden som rapporterats i andra studier. En jämförelse av övriga mätvärden med tidigare litteratur är dock svår på grund av brist på liknande studier. Trots detta har vissa studier som jämfört IMU och MoCap vid uppskattning av PO under countermovement jumps rapporterat en överskattning på cirka 25%, vilket ligger nära den råa överskattningen på cirka 33% i denna studie.

Studien visar att den mest tillförlitliga metoden för att uppskatta PO med IMU är FA. Dessutom visar resultaten att även om en effektmätare med tre IMU:er ger lovande resultat, så ger konfigurationen med endast två sensorer (på bröstkorg och handled) något bättre skattningar.

**Nyckelord**

IMU, MoCap, Functional Fitness, CrossFit, effektmätare, validation

*This page intentionally left blank.*

# Acknowledgments

First and foremost, I would like to thank my supervisor Mariah for having given me the opportunity of doing such an interesting study: I learned a lot during these months, and it has been very interesting and challenging to try to develop something completely new. I would also like to thank all the athletes who took part in the data collection, your participation meant a lot. I'm also grateful to the staff at the Promobilia MoveAbility Lab in KTH for lending me their equipment and for their valuable assistance.

Lastly, many thanks to my family, that has always supported me in this experience abroad, and to all my friends, in Italy and in Sweden, for having shared with me this incredible journey and made it nicer with their presence.

Stockholm, August 2025
Nicolò Rubele

*This page intentionally left blank.*

# CONTENTS

*This page intentionally left blank.*

# 1 Introduction

Power meters have shown, since their first development in the 1980s, that they are useful tools that can help in developing more efficient training programs for athletes and in reducing the risk of injury [1]. Despite these benefits, their adoption is still mainly in endurance sport [2], especially cycling, while their presence in gym contexts is limited to machinery like the bicycle, the rower or similar. One sport that could benefit a lot from the usage of a power meter could be functional fitness: indeed, due to the high variability in the exercises prescribed for each session and the difficulties of estimating the impact of each exercise on the whole program [3], a reliable method that can assess the performance of the athletes, helping in developing a better understanding of how their bodies are reacting to the stimuli provided them, can help in booster athletes performance keeping, at the same time, injury risk at bay. So, this work consists of one article that shows the development of an inertial measurement unit (IMU)-based power meter designed for functional fitness athletes and its validation against a Motion capture system (MoCap) coupled with 4 force plates.

This work provides a broader context, discusses methodological choices in detail, and reflects on the implications of the findings for on-field application and future research.

# 2 Background

Since 1980s, when the first power meter was developed [1], power as a sport metric has gained more and more attention by several disciplines. The early adopters of this technology were cycling teams [1] while, in more recent years, thanks to the evolution of IMU sensors, easy-to-use tool to monitor power allowed the possibility to monitor this parameter in an increasing number of other sports, such as running [2]. The underlying reason behind its wide adoption it that the power meter provides useful insights for assessing performance with reliability. Nevertheless, while this metric, and the relative device to measure it, is well established in the endurance sports, in the gym context it is available just for those machineries that allow for endurance training, such as the rowing machine and the bicycle. In all the other gym disciplines there is not a consensus about which metrics to use to assess performance, and this is particularly true for functional fitness. [3]

In Functional fitness, of which CrossFit® is the most popular example, the main problems in trying to assess athletes' performance are two, according to Mangine and Seay [3]:

1. Different parts of the training sessions are measured with different methods that output metrics with different units;
2. The same exercise, put in different parts of the training session, can produce different stresses in the body that can be difficult to quantify using separate metrics for every training component.

There have been many attempts to try to overcome the problems before mentioned and trying to quantify and compare different workout strategies in functional fitness but, as stated earlier, a consensus has been not achieved.

For this reason, it seems promising to try to introduce a power meter, and the performance assessment that it enable, into the functional fitness world. The main barrier to its adoption, though, is the lack of such a device that can work specifically for functional fitness. Indeed, the great variety of movements performed by the athletes and the fact that they may have breaks in the middle of the movements pose difficulties in an effective estimation of the power output. This study aims to bridge the gap between the necessity of performance monitoring in functional fitness and the lack of a power meter that can help in this. To do so, an IMU-based power meter designed for functional fitness is developed and validated against MoCap.

# 3 Research Methodology

This section, with all its subsections, will go deeper into the methodology that was followed during the study. In particular, the focus will be on the design of the research questions, on how the literature review that has been conducted and on the methodology approach.

## 3.1 Design of Research Questions

The initial interest in this research was the idea of applying power output, a metric wildly used in many sports to assess athletes, to functional fitness. To start to dig into this topic, preliminary research in the literature was performed to see if there were already previous studies that attempted to apply power output in order to assess performance in functional fitness athletes. The results provided no results other than some articles highlighting the necessity of power monitoring, to be able to improve performance, and some initial estimation of power output in some specific movements such as clean and jerks and snatches. These first estimations, though, were performed analyzing video recordings of the movements or with rough calculations based on simple physical laws. So, there seemed to be a lack of reliable methods to provide power output estimations almost in real time, in order to give athletes useful feedback. Contextually, during the power estimation literature review, many articles about the estimation of power during countermovement jump with the use of IMUs appeared. So, the following research questions were formulated:

- How is it possible to sum up all the suggestions available in literature in order to develop an IMU-based power meter?

- How accurate and reliable can the power estimation of a device like this be?

These questions seemed to be the first step to try to bridge the current gaps in literature. The decision to investigate both the possible technical approaches to develop the power meter and the validity of such a device contributed to a real advancement in the possibility of athletes monitoring, not only producing a tool that works in theory, but also proving how reliable the provided power estimations are.

Other research questions were taken into consideration. For example, a question could have been if it was feasible to develop a power meter for functional fitness with other technologies rather than IMUs, or which technology was the best solution to develop such a device. These questions were excluded because the idea was to have a working prototype of the power meter after the study and, to mainly focus on the research and development of several solutions with different technologies would have taken too much time to be able, then, to perform a good validation study with all of them. So, considering the promising results provided by IMUs in the countermovement jump power estimation and in other sports, the decision was to focus directly on them. Other questions, in addition, were not taken into consideration because they were referring to a later stage of the path toward bridging the literature gap. For example, a question could have been if the use of a power meter in functional fitness is really effective and, in that case, how the use of a power meter would have benefitted functional fitness athletes both from a performance and from an injury risk reduction perspective, so an assessment of the real differences that can be made with a device like this in the design of training programs.

## 3.2 Introduction Framework

To develop the Introduction section of the study, a literature search was employed to ensure the inclusion of relevant and high-quality literature. Searches were conducted in **PubMed** and **Google Scholar**, as these databases cover a wide range of medical, technical, and interdisciplinary research.

The following keyword combinations were applied:

- ("mechanical power" OR "power output") AND ("functional fitness" OR "CrossFit")
- ("power output" AND "IMU") AND (“functional fitness” OR “CrossFit”)
- ("power output" AND "IMU" AND “sport”)

- (“power estimation” OR “power output estimation”) AND “sport”
- (“IMU” AND “placement” AND “optimization” AND “sport”)

The search yielded 735 articles considering the following inclusion criteria applied:

- Articles published after 2015 to ensure the latest advancement in the technology for the IMU and the most recent researches in the functional fitness field.
- Peer-reviewed journal articles or conference proceedings.
- Studies evaluating strategies to monitor power output or performance in functional fitness or other sports, studies evaluating the use of IMU in functional fitness or other sports and studies evaluating optimal IMU placement in sport.

Some more articles were found reading the bibliography of articles that were particularly centered on one of the topics of interest. After reviewing titles and abstracts and removing duplicates, 51 articles were selected for full-text review, and ultimately, 27 articles were included in the final background section.

During this process, it was observed that there were some studies trying to estimate power output in functional fitness or trying to find methods to assess athletes’ performance but there were no studies providing reliable tools that could help in assessing athletes regarding this aspect. This confirmed the need for an effective and reliable power meter to be used in functional fitness.

## 3.3 Methodology Approach

A **quantitative-method approach** was selected to address the research questions effectively. This approach used the quantitative data reported in literature to select the strategy to develop the power meter and then the validation part was carried out with a quantitative analysis of the newly developed power meter to determine its accuracy, precision and agreement with MoCap.

### 3.3.1 Power output estimation strategies

To address the first research question, data available in literature were used and combined to come up with the best strategies to estimate power.

This method was chosen because it provided a solid foundation for developing a device that could estimate power based on already tested conceptual models. An alternative approach could have been just focusing on biomechanics and physics laws to find a completely new estimation strategy. This approach, though, was discarded as it would have required a longer time only to generate a strategy that was not proven to work in practical terms, requiring an additional step to verify the theory behind the estimation strategy first.

#### 3.3.2 Power meter performance evaluation

To address the second question, power meter performance in estimating power output was evaluated recording functional fitness movements in a laboratory both with the IMUs and the MoCap, used as gold standard, and then the power output obtained from the two systems was compared in order to determine the following aspects:

- Accuracy of the estimation: the power meter's ability to produces results close to the gold standard one;
- Precision of the estimation: the power meter's ability to produce results that are close between them;
- Agreement of the estimation: the power meter's ability to be consistent with the values produced by the gold standard system.

This method was chosen because it provided good quantitative metrics to quantify the performance of the new power meter compared with the gold standard instrument for computing the power output, the MoCap. An alternative approach could have been to validate the device directly in a gym, instead of a laboratory, but no reliable ground truth systems were found to be suitable for a proper validation study.

#### 3.3.3 Alignment with Research Questions

The chosen methodology was designed to directly align with the research questions:

- Literature data combination addressed the first research question by providing a proven model to estimate power;
- Power meter evaluation against a gold standard system in a laboratory allowed to answer to the second research question by providing quantitative results to assess its performance in estimating power in a high controlled environment.

By combining these methods, it was possible to have a solid and verified estimation model to be used for the power meter and then to validate the device with limited amount of possible interfering factors.

One limitation of the present methodology is that the movements were recorded in a laboratory, so outside of the usual environment of the athletes. Moreover, the athletes had 70 reflective markers on their body to record the movement with the MoCap system, and they were forced to perform all the movements on 4 force plates, thus with possible modification of athletes' technique during the test. Future studies could try to find reliable systems that allow the test to be carried out in a gym, where athletes are free to move without the constraints of the laboratory. Additionally, the sample size for the test was limited to 7 participants, which may not be representative of all athletes, thus limiting the generalizability of the findings. Expanding the study to include more athletes and, maybe, with different levels of ability could provide a broader perspective.

# 4 Summary of Article

The following sections will highlight the main findings of the study, reporting shortly the results and the discussion that can be found in their complete version in the study paper.

## 4.1 Results

The results of this study showed that two approaches are possible for estimating power output from IMU data: the first one, the force approach (FA), is based on the definition of power (P) as product of force (F) and velocity (v), according to the formula $P = F * v$, while the second one, the energy approach (EA), is based on the definition of power as work performed, or energy transferred, over time by the athlete. This second approach computes power (P) as the difference in mechanical energy over time, according to the formula $P = \frac{\Delta(E_{kin}+E_{pot})}{\Delta t}$, where $E_{kin} = \frac{1}{2} * m * v^2$ is the kinetic energy and $E_{pot} = m * g * h$ is the potential energy, the two components of mechanical energy.

When comparing the results of the power output estimated with these two methods with the power output computed using MoCap, both the approaches showed very high to almost perfect correlation using Pearson correlation coefficient. FA showed a slightly better correlation compared to EA (r = 0.908 for FA vs r = 0.894 for EA, in the configuration with 3 sensors). This is in line with other studies that compared the correlation between IMU and MoCap: indeed, Jimenez-Olmedo et al. [4] reported a correlation of 0.847 for the pelvis level, while other studies [5], [6], [7] reported a correlation between 0.72 and 0.95 for IMU worn at torso level.

When analyzing the best configuration between the one with 3 sensors (ankle, chest and wrist), the ones with 2 sensors, one with sensors worn at ankle and chest (2a) and one with sensors worn at chest and wrist (2w), and the configuration with just one sensor at the chest, the best correlation was obtained for the configuration 2w, both for the force approach and the energy approach.

Investigating accuracy and precision for the two approaches and their agreement with the MoCap, two different scenarios were analyzed: in the first one the data outputted by the newly developed power meter were compared directly with the MoCap (raw estimation), while in the second scenario a correction to the estimation was applied before the comparison (corrected estimation). The correction was done based on the linear regression equations computed between the power meter power estimations and the MoCap estimations.

Considering the raw estimations, both the approaches overestimated the power output in all the metrics analyzed (average positive and negative power, total positive and negative work and peak positive and negative power) by around 33%. The energy approach outperformed the force approach in almost all the metrics in agreement and accuracy (mean percentage error: 34.2% FA vs 28.1% EA, excluding the only metric better estimated by the FA, peak positive power, which has an accuracy of 31.61% FA vs 59.02% EA), while the precision is the same for both methods. No similar studies compared power output estimation for functional fitness movements, but some studies compared the mean power output for countermovement jump estimated with IMU and with force plate as gold standard. Despite the movements are different, it was possible to compare their percentage error with the one found in this study to have a general idea of the power meter performance. Rantalainen et al. [6] reported a percentage error of 25.8%, smaller than the 36.14% found in this study with FA and the 31.54% with EA. But, as already mentioned, the movements are different, so a direct comparison is difficult.

After the correction, precision, accuracy and agreement improved for all the metrics and with both approaches, but still showing a big difference in the estimation of peak positive power: the mean percentage error becomes -10.2% for FA and -10.7% (excluding the peak positive power, which accuracy is -2.55% for FA and 22.40% for EA). As it is possible to see from the percentage error, after the correction both the approaches underestimate the MoCap data except for the peak positive power for energy approach.

Finally, considering the best sensor configuration, the 2w estimated using FA showed to be the one providing the best results (mean percentage error: -8.9% for configuration 3 vs -7.3% for configuration 2w).

## 4.2 Discussion

Power output estimated with the newly developed power meter showed a very high to almost perfect correlation with the power output computed from MoCap when comparing the power curves for the whole duration of each movement. This indicated that the new IMU-based power meter can be a valid tool to be used in estimating power and this was supported by the analysis of some metrics derived from the whole curve, as the one already mentioned in the results section. Indeed, considering that, at the moment, there is nothing that can help coaches and athletes to have indications for the development of training programs and to assess performance in athletes, the power meter developed in this study can give a lot of useful information and can help in improving performance and in reducing, at the same time, the risk of injury. Of course, it must be taken into account that the best configuration found in this study (the one with 2 sensors applied at chest and wrist and with power estimated with FA) has still a 7% bias toward the underestimation in its power estimation but, for now, this value is considered acceptable to be used in a sport context. For the future some other strategies can be implemented and tested to try to further reduce the bias and increase the accuracy of the power meter.

Thanks to these findings, it is now possible to bridge the current gap existing between endurance sports, where power meters have been used for many years to assess performance and design highly individualized training plans, and functional fitness, where nowadays there are still debates on how to properly measure athletes' performance. From now on, it will be possible to estimate power produced during the different parts of a functional fitness training program, in order to have a broader view over each training session.

# 5 Fulfillment of Learning Outcomes

This section will show how the common goals and assessment criteria of the Degree project course have been achieved with the thesis work presented in the study paper.

## 5.1 Scientific Knowledge and Methodology

This study shows a deep understanding of the fields of power output in sport, IMU-based power estimation and the capacity of developing and validating new devices. A thorough literature review has been conducted to understand state-of-the-art technologies and methodologies for power estimation in biomechanical and sport contexts and how IMU can help in this process. Two main approaches in power estimation, such as the computation of power according to the formula $P = F * v$ (force approach), or the computation of power as the rate of mechanical energy transformed over time (energy approach), were integrated in the development of the IMU-based power meter.

For the methodology part, the study used a quantitative method to validate the new power meter analyzing its correlation, precision, accuracy and agreement with a gold standard technology (motion capture). This is the currently most used approach to validate new measurement systems. More in-details explanations of the methodology and the choices behind it can be found in the methodology section (Section 3).

## 5.2 Systematic Information Gathering

The literature review process done as basis of this study was conducted using PubMed and Google Scholar and with specific keyworks combined together in order to find the most relevant articles. To refine the search strategy, a filter was used to search for articles published in the last 10 years and to search only for scientific studies articles. Moreover, reading the bibliography of the articles proved to be particularly centered on the topics of interest, additional articles were included in the review process.

The review revealed a lack of studies about reliable and effective ways of measuring performance in functional fitness, helping in focusing the aim of the present study. To bridge this gap a new power meter was developed and a validation study was conducted. The Background section of this article (Section 2), the Methodology part concerning the literature review (Section 3.2) and the Introduction section of the study show how the information was gathered and integrated together.

## 5.3 Handling Complex Questions

The present study deals with two complex questions: the first one is how it is possible to estimate power output in functional fitness using an IMU system, while the second one is the precision and accuracy of such a system. For both the questions there was limited information available in literature, as, presently, there were not many articles analyzing the different approaches available for power estimation and no one studying the performance of a power meter device in functional fitness. Expanding to the use of power meters in other non-endurance sports, the information remained limited as there were just some power output estimations done with an IMU during countermovement jumps. Despite this reduced number of studies, the need for a tool to assess performance in functional fitness was real, with several articles trying to find alternative metrics to correctly estimate performance or ways to estimate power output. These complexities are analyzed and discussed in the "Discussion" section of the study.

## 5.4 Planning and Execution

The research followed a structured timeline, with predefined milestones for literature review, algorithm development for power estimation, data collection, and quantitative analysis. The initial research plan outlined in the thesis proposal was followed, with minor modifications agreed upon with the supervisor when the data collection phase was prolonged in respect to the period that was in the plan. The study methodology is explained in detail in the “Material and methods” section of the study. No big modifications required to be reported in the study to maintain transparency.

## 5.5 Clear Communication and Argumentation

The study article follows the typical scientific study structure, with an introduction section, the material and methods part, the results and the discussion parts and, finally, the conclusion. This allows for a logical, clear and well-structured progression of the topics, allowing the reader to have the context to be able to understand the study results and to grab the implications of what has been discussed. Moreover, all the key claims presented a strong justification and were connected, when possible, with similar findings in literature. This can be seen in the Discussion section of the study, where the main findings of the study were presented and discussed in a broader context.

No artificial intelligence tools were used in any aspect of the writing phase of the present work.

## 5.6 Scientific, Social, and Ethical Assessments

The study considered the ethical, social and scientific implications of developing a new power meter to be used by athletes and coaches in order to assess performance and enabling a more accurate training program design. The ultimate goal is to improve athletes’ performance and reduce the risk of injury. For this reason, it is important to have a device that is scientifically validated in order to avoid the risk of providing wrong results that could potentially lead to injuries in athletes. Indeed, the risk is that wrong power output estimations can suggest to coaches and athletes some training sessions that could be harmful to athletes resulting in an injury or, less severe, in a reduction of performance. This can be seen both as an ethical and social problem, as the athletes need to suspend training, spend money on healthcare to recover from the injury, without considering their physical and psychological suffering. These considerations informed some reflections that were made, for example, in the Discussion section of this article (Section 4.2) and of the study.

## 5.7 Research and Development Readiness

This study highlights the capacity and readiness to work in the research and development field, especially in the sport sector and using sensors. Indeed, this work consisted in the assessment of the state-of-the-art with a literature review, in the the design and development of a new power meter, integrating all the information found in literature, and in a validation study to assess the performance of the new device compared to the gold standard. Moreover, in the Discussion section of the study, it is explained how the findings of this study contribute to the development of functional fitness, helping the coaches to unleash the potential of their athletes, but also to the development of new power meters for non-endurance sports.

# References

[1] L. Passfield, J. Hopker, S. Jobson, D. Friel and M. Zabala, "Knowledge is power: Issue of measuring training and performance in cycling," *Journal of Sports Science,* 2016.

[2] D. Jaén-Carrillo, L. Roche-Seruendo, A. Carton-Llorente, R. Ramirez-Campillo and F. Garcia-Pinillos, "Mechanical power in endurance running: A scoping rview on sensort for power output estimation during running," *Sensors,* 2020.

[3] G. Mangine and T. Seay, "Quantifying CrossFit: Potential solutions for monitoring multimodal workloads and identifying training targets," *Frontiers,* 2022.

[4] J. Jimenez-Olmedo, B. Pueo, J. Mossi and L. Villalon-Gash, "Concurrent Validity of the Inertial Measurement Unit Vmaxpro in Vertical Jump Estimation," *Applied Sciences,* 2023.

[5] R. Spangler, T. Rantalainen, P. Gastin and D. Wundersitz, "Inertial Sensors are a Valid Tool to Detect and Consistently Quantify Jumping," *International Journal of Sports Medicine,* 2018.

[6] T. Rantalainen, P. Gastin, R. Spangler and D. Wundersitz, "Concurrent validity and reliability of torso-worn inertial measurement unit for jump power and height estimation," *Journal of Sports Sciences,* 2018.

[7] J. Pino-Ortega, J. Garcia-Rubio and S. Ibanez, "Validity and reliability of the WIMU inertial device for assessment of the vertical jump," *PeerJ,* 2018.

*Article*

# Development and Validation of a Power Meter for Functional Fitness Athletes

**Nicolò Rubele [1]*, Mariah Marques Sabioni [1]**

[1] Royal Institute of Technology (KTH), Stockholm, 11426 Sweden
* Correspondence: rubele@kth.se

**Abstract:**

Power meters have been available in endurance sport since the 1980s but in gym context, especially for functional fitness (FF), it is still not possible to measure the power output (PO) of athletes, thus making it difficult to understand how different training sessions can affect athletes' bodies, both in terms of performance and fatigue. For this reason, the aim of this study is to develop a new inertial measurement unit (IMU)-based power meter to be used in FF contexts and to validate it against a combination of a motion capture system (MoCap) and 4 force plates. Seven well trained FF athletes took part in this study. They were asked to perform some FF movements while being recorded by 3 IMUs placed on them (chest, wrist and ankle) and by the gold standard system. The new power meter estimated the PO with different combination of the 3 IMUs and with two different approaches: the force approach (FA), based on the formula P=F*v, and the energy approach (EA), based on mechanical energy. Both FA and EA showed very high to almost perfect correlation with the gold standard, with the best correlations provided by FA, when comparing the full PO curve. Analyzing average positive and negative power, total positive and negative work and peak positive and negative power, FA showed better estimations despite a general underestimation of around 10% for all the metrics analyzed and with a RMSE around 25% (after the correction of the raw estimations, while raw PO overestimates the MoCap by around 33%). This study showed that the best method to estimate PO with IMUs is FA. Moreover, even if the 3-IMU power meter showed promising results, the configuration with just two sensors (at chest and wrist) provided slightly better estimations.

**Keywords:** IMU, MoCap, Functional Fitness, CrossFit, power meter, validation

## 1. Introduction

Since the 1980s, when the first power meter was developed **[1]**, coaches, sport scientists and athletes have had the possibility to use a new and very powerful tool to study performance: mechanical power output. The early adopters of this technology were cycling teams **[1]** which use it to get power profile, training load, and performance assessments of athletes and also to create tailor-made training zones for every single athlete **[2]**, in order to develop their potentialities in the best possible way.

To do this, a concept widely used in cycling is the Functional Threshold Power (FTP): as reported by Allen et al. **[3]**, FTP is the highest power output that an athlete can produce in a quasi-steady state for approximately one hour. In order to find this threshold for every athlete, many different tests are available. **[3]** Once obtained this number, it is possible to find different training zones based on the percentage of FTP and specifically designed to address some specific aspects of the training, such as endurance training (56%-75% of FTP), lactate threshold training (91%-105% of FTP) or anaerobic training (21%-150% of FTP) **[3]**. The use of power meter, in all of this, not only helps in defining FTP, but also in monitoring performance changes and also to train in the desired training zone, thus boosting performance **[3]**.

Thanks to all the benefits that the use of power meters give, it is not a surprise that cycling has been an inspiration to also provide mechanical power estimation in other cyclical sports **[2]** and in running **[4]**. This adoption is seen as a big step forward and as reported by Jaén-Carrillo et al. **[4]** (p. 1): "[power meter] might also change the way runners compete and train".

Nevertheless, while this metric, and power meters, are well established in endurance sports, in the gym context it is available just for those machineries that allow for endurance training, such as the rowing machine and the bicycle. In all other gym disciplines, where the athletes perform unconstrained exercises, not only is it not possible to measure power output, but also there is not a consensus about which metrics to use, and this is particularly true for functional fitness **[5]**.

Functional fitness, of which CrossFit® is the most popular example, is "a training style [or program] that incorporates a variety of functional movements, performed at high-intensity […], and designed to improve parameters of general physical fitness […] and performance" **[6]** (p. 2). In this definition two points are worth mentioning: the first one is the focus on high-intensity movements. In CrossFit® methodology **[7]** intensity level is defined by power output, so the role of power in this program is very important. However, there is a lack of explanation on how to measure it. The second point to mention from the definition is the explicit prescription of "constantly varied […] movement" **[8]** (p. 1) that makes performance measurement very difficult in this discipline. The main problems here are two, according to Mangine and Seay **[5]**:

1. different parts of the training sessions are measured with different methods that output metrics with different units;
2. the same exercise, put in different parts of the training session, can produce different stresses in the body that can be difficult to quantify using separate metrics for every training component.

There have been many attempts to try to find some values that can help in estimating performance **[9]**, **[10]** or attempts in finding strategies to keep track of the different training components **[11]**, **[12]**, **[13]**, **[14]**, **[15]** for overcoming the problems mentioned before but no one of them seems to really solve the problem.

However, the Level 1 CrossFit Training Guide **[16]** provides a way to estimate the workload and the power for some workouts and, although this is just an approximation, "it is the first instance where the contribution of individual workout components to the overall workload were considered" **[5]** (p. 8).

A possible way to improve this promising strategy, by overcoming the approximations in the method used in **[16]**, could be to track athletes' movements using one or more inertial measurement units (IMUs), as it is actually done in running **[4]**, providing in this way more precise results.

The aim of this study is, then, to develop and validate a power meter based on IMUs and specifically designed to be used by functional fitness athletes, to provide coaches and athletes with useful insights into their performance and indication of where to improve.

Before going deeper into the validation study, the necessary theoretical background, about how to perform power estimation and some technologies that allow to do that, will be provided.

*1.1 Power and power estimation*

Mechanical power is defined as "the rate at which the athlete does work or transfer energy to complete a movement task" **[17]** (p. 2) and, as mentioned earlier, the interest in this metric is due to several information that it is possible to draw from it, according to de Vette et al. **[2]**:

- it can be used as a performance measure;
- it is an objective way of measuring the external load sustained by an athlete, given that it takes into account several environmental factors that can influence a training session or a competition (e.g. wind velocity);
- it can be used to assess fitness and fatigue in athletes, helping to prevent overtraining and in periodize training.

Deriving from the definition, power in physics is defined as $\boldsymbol{W/t}$, where W is the work performed and t is the time under which the work is done. But the power associated with a force can also be defined as the cross product $\boldsymbol{F \cdot v}$, where F is the force applied to a body and v is the velocity at which the body moves due to that force.

From this last definition, considering all the forces acting on an athlete's body part *i* and the velocity under which part *i* moves, it is possible to derive the power equation for part *i* in its translational part by multiplying Newton equation with the velocity:

$$\sum_i \vec{\boldsymbol{F}}_{\boldsymbol{i}} \cdot \vec{\boldsymbol{v}}_{\boldsymbol{i}} = \boldsymbol{m_i} \cdot \vec{\boldsymbol{a}}_{\boldsymbol{i}} \cdot \vec{\boldsymbol{v}}_{\boldsymbol{i}} \qquad (1)$$

where $\boldsymbol{F_i}$ are all the forces applied to the body segment *i*, $\boldsymbol{m_i}$ is the mass of the segment *i* and $\boldsymbol{a_i}$ and $\boldsymbol{v_i}$ are the related segment linear acceleration and velocity due to the forces.

Similarly, it is possible to write the power equation in its rotational part, due to the moments applied to the different body parts *i* of the athlete, using the Euler equation and multiplying it with the rotational velocity of the body part *i*:

$$\sum_i \vec{\boldsymbol{M}}_{\boldsymbol{i}} \cdot \vec{\boldsymbol{\omega}}_{\boldsymbol{i}} = \frac{\boldsymbol{d}}{\boldsymbol{dt}}(\boldsymbol{I_i} \cdot \vec{\boldsymbol{\omega}}_{\boldsymbol{i}}) \cdot \vec{\boldsymbol{\omega}}_{\boldsymbol{i}} \qquad (2)$$

where $\boldsymbol{M_i}$ are all the moments applied to segment *i*, $\boldsymbol{I_i}$ is the inertia of the segment *i*, and $\boldsymbol{\omega_i}$ is the related segment angular velocity due to the moments.

As reported by many authors **[2]**, **[17]**, in an athlete we can have four different force (and corresponding moment) origins, namely joint forces $\boldsymbol{F_j}$, gravitational forces $\boldsymbol{F_g}$, external forces $\boldsymbol{F_e}$, and frictional forces $\boldsymbol{F_f}$. Summing up the two components in (1) and (2) and separating the origin of all the different acting forces and moments, we can finally write the final power equation:

$$\boldsymbol{P_j} = \boldsymbol{P_k} + \boldsymbol{P_f} - \boldsymbol{P_g} - \boldsymbol{P_e} \tag{3}$$

where $\boldsymbol{P_j}$ is joint power, $\boldsymbol{P_k}$ is kinetic power, $\boldsymbol{P_f}$ is frictional power, $\boldsymbol{P_g}$ is gravitational power, and $\boldsymbol{P_e}$ is environmental (i.e. external forces and moments) power. This means that the power generated by an athlete (joint power) is used to generate movement (kinetic power) and to overcome resistive forces, such as friction, gravity and external forces (frictional, gravitational and environmental powers).

Solving this power equation while treating athlete's body as a chain of a number of linked rigid bodies allows for mechanical power estimation **[2]**. Equation (3) clearly shows that there are 2 different methods to estimate the mechanical power output of an athlete: through the calculation of the joint power, determined multiplying the joint torque and angular velocity, or through the calculation of all the components on the right side of the equation (3). As reported by van der Kruk et al. **[17]**, both these approaches are suitable for estimating power output.

To estimate the power of the right side, two approaches are possible **[17]**:

1. using the instantaneous power and solving the equation (3) for every timeframe (e.g. using force and velocity derived from a force plate **[18]**);
2. determining the change of kinetic and potential energy of a system. This approach was already performed by Garhammer **[19]** in 1993.

These two methods can provide very good estimations but, considering that usually not all the terms on the right side are estimated, the joint power approach could be seen as a "gold standard" in the power calculation. As a consequence, the two methods to evaluate the right side of equation (3) can lead to a simplified model, neglecting some components. This model, though, is very useful for a field-based estimation of the power, since both the joint power calculation and the complete calculations of all the other components in equation (3) can be very laborious **[2]**.

Indeed, one way to be able to estimate the joint power is to measure, in a laboratory context, all the biomechanics variables through motion capture systems, 3D cameras, multiple inertial measurement units (IMUs) throughout the whole body (ideally one per each segment) or cameras paired with neural network models that are able to recreate the 3D movement of the athlete **[20]**. But most of the applications of mechanical power require constant monitoring and in an environment that is the one where the athletes usually perform their activities **[2]**. So, the laboratory environment seems not indicated to be able to get all the benefit deriving from the adoption of a power meter.

To be able to obtain valid data to be used with the right side of the equation (3), it is required to use wearable devices, such as IMUs, to measure body segment kinematics **[2]** in a way that they do not impair the ability of an athlete to perform his/her usual activity (i.e. to not have an IMU per segment). Moreover, IMUs can speed up the process of kinetic and potential energy estimation, instead of performing a very long video analysis, as the one conducted by Garhammer **[19]**.

IMUs are small and lightweight devices that are composed of accelerometers, gyroscopes, and magnetometers **[21]** and a limited number of IMUs placed on the athlete seems the best solution to be able to estimate the mechanical power output. With this equipment, they can collect data about linear acceleration, angular velocity, and local magnetic field of the segment where they are attached to. The output of an IMU can be used to determine the kinematics of the segment they are attached to (i.e. its orientation and angular velocity) but they are useful to estimate external forces as well, making them the perfect "standalone approach for in-field mechanical power estimations" **[2]** (p. 2).

Despite the fact that IMUs are small and lightweight, it is not feasible to think that an athlete can wear a huge number of these sensors during his/her normal activity. For this reason, it is important to think well about how many sensors to use and where to put them to obtain a good trade-off between the need for good and reliable data, required by coaches and sport scientists, and a movement as natural as possible avoiding that IMUs interfere with athletes' performance.

### *1.2 IMU placement*

There has been a boom in the last decade of researches about the use of IMU sensors in sports **[22]** thanks to their characteristics highlighted in the previous paragraph. Among these studies, many focused on the best placement of IMUs in order to get the best kinematics characteristics of the subjects analyzed, for the whole body **[23]**, **[24]** or specifically for upper **[25]** or lower limbs **[26]**. Despite a review performed in 2010 concluded that the validity of the IMUs depends on the joint studied and movement performed **[27]**, several articles have been analyzed to try to find the most promising locations to compute mechanical power output. The requirements behind this research were two:

1. to find locations that could represent both the limbs and the trunk, in order to take into account all the different body parts that can contribute to the total power output;
2. to avoid the usage of too many sensors, otherwise they could impair a natural athletic performance.

Starting from a systematic review conducted in 2019 by Poitras et al. **[23]**, it was possible to analyze pros and cons of every joint and the most promising locations were the wrist, the trunk, the pelvis, and the ankle.

All these locations showed both good validity and reliability. The ankle was also backed up by the study conducted by Rahn et al. **[24]**, while the wrist position was suggested by Walmsley et al. **[25]** as well. Moreover, a recent study, in which IMU sensors were used in order to assess countermovement jump and squat jump tests, found out that the performance of sensors placed on the chest and on the pelvis is very similar. Despite this, on average, the chest position demonstrated greater reliability. **[28]** The location on the chest can be preferred for two other reasons:

1. the sensor used in this study (Movesense Flash OP174) is also able to monitor heart rate, so the placement on the chest allows us to obtain this information too;
2. in many sport situations, rules expressly say that IMU must be worn on the torso, as reported by Mohammadin et al. **[29]**

So, the locations selected for the present study are the wrist, to monitor the movements of the upper limbs, the ankle, to monitor the lower limbs, and the chest to monitor the trunk.

To avoid redundancies and the usage of too many sensors, the IMUs will be placed only on one upper and one lower limb.

## 2. Materials and Methods

*2.1 Participants*

A total of 7 athletes (age: 32.43 ± 1.99, height: 173.00 ± 11.50 cm, weight: 73.57 ± 12.90 kg, experience in functional fitness: 6.71 ± 2.75 years) from different Functional Fitness gyms in the Stockholm area took part in this study. Before starting the study, all of them signed an informed consent in which there was an explanation of how their data would have been treated. Moreover, they received an explanation of the study, the movements to perform and the weights they were supposed to use for each movement.

The following criteria were applied to select the participants:

- inclusion criteria: being at least an intermediate functional fitness athlete with at least 2 years of experience; being an athlete training at least 3 times per week; being proficient in all the movements that were in the study; having a 1RM (the maximum weight that can be lift for one repetition) for the snatch of at least 60 kg;

- exclusion criteria: having sustained an injury in the previous 6 months or presenting other pathologies for which performing physical activity was a contraindication.

*2.2 Study design and instruments*

The present study was structured to be an observational study to determine the validity of a newly developed IMU-based power meter. The data collection for this study was conducted in the Promobilia MoveAbility Lab, the biomechanical laboratory at the Royal Institute of Technology (KTH) in Stockholm, Sweden, during the spring and the summer 2025. The laboratory is provided with a Motion capture system (Vicon) consisting of 10 Vicon cameras "Vicon Vantage 16" (16 MP at 120 Hz), 2 video cameras "Vicon Vue" (720p at 120 Hz) and 4 force plates "AMTI BMS400600".

For the present study several acquiring technologies and software were used in order to have robust data to create and validate the power meter:

- The already mentioned Vicon system, sampling at 100 Hz both with Vicon cameras and video cameras and at 1000 Hz with the force plates. The Vicon cameras were used to detect reflective markers applied to the athletes, while the video cameras were used to have a visual recording of the trial. These two video cameras were placed one in front of the athletes and one on their left side. The force plates were used to measure ground reaction force (GRF) during the movements. All these instruments were synchronized between them;
- 3 IMUs (Movesense Flash OP174, with heart rate recording integrated) sampling at 52 Hz, to record the data to develop the power meter, and connected via Kaasa app to an iPad to control them;
- Vicon Nexus 2.15, to post-process the data coming from the motion capture system;

- OpenSim 4.5, to be able to compute joint moments and displacements from the motion capture data;
- Visual Studio Code and Python 3.10.7 to create the scripts to analyze the IMU data and create the power meter pipeline.

*2.3 Experimental procedure*

To ensure optimal data collection with the motion capture system, the system was turned on one hour before the athletes arrived and a calibration procedure was performed. Then, a protocol was followed during the data collection to provide uniformity in the process. The protocol consisted in the explanation of the agenda to the athletes, the collection of their informed consent, some questions regarding their experience in functional fitness and with the movements required for this study, and a warm-up phase. During this phase, the three IMUs to be used to collect data were placed on a flat surface and then some seconds were recorded in order to perform a static calibration. After the warm-up, the three IMUs were placed on the chest, on the wrist and on the ankle (Figure 1).


*Figure 1: Sensor placement on the athletes. The sensors used in this study are the one indicated with the red arrows (the chest one is not visible as it is under the sport bra)*

67 reflective markers, to be used in combination with the motion capture system, have also been placed on the subjects following a modified CGM2 schema (see the small white dots in Figure 1 and Figure A1 in Appendix for a representation of the modified CGM2 schema used).

Once the setup was done, each athlete was asked to perform 5 repetitions of the following functional fitness movements (Figure 2) while being recorded both with the motion capture system and the IMUs:

- burpees, with 3 different techniques:

   - double step, in which the athlete steps down in the plank position before the pushup and then steps up after the pushup;
   - step up, in which the athlete jumps down in the plank position but returns from it stepping up;
   - jump up, in which the athlete performs a jump both going down in the plank position and returning from it;
- power clean and push jerks;
- forward lunges (to be performed without any additional weight, with a 20 kg barbell and with a 30 kg barbell);
- power barbell-snatches;
- thrusters.

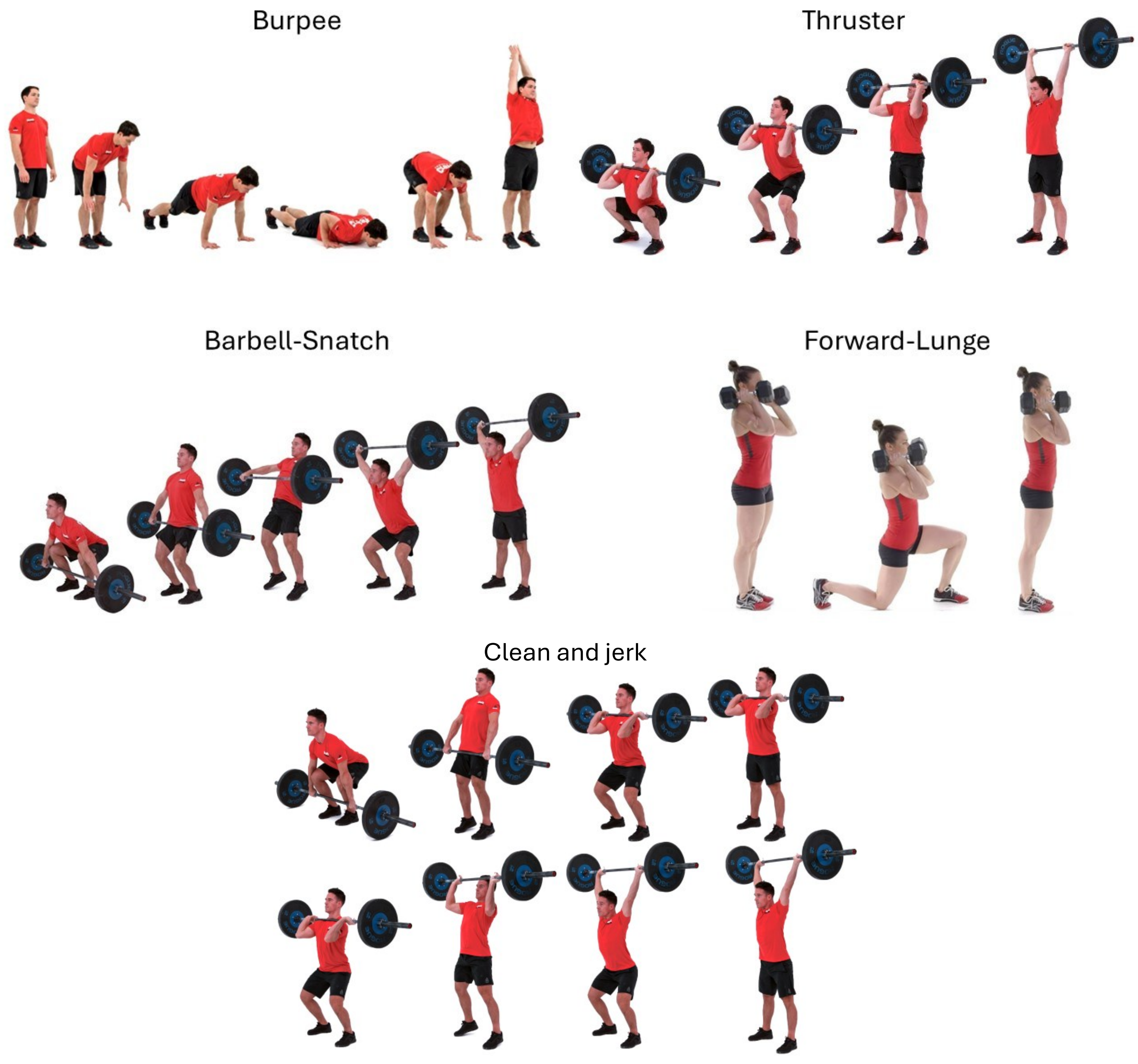



*Figure 2: Functional fitness movements performed in the study, taken from crossfit.com*

The only movement with a different number of repetitions was the forward lunge because of the disposition of the force plates on the floor. So, the athletes were asked to perform two trials with each external weight, producing 4 valid repetitions per weight. Only those 4 repetitions were the ones analyzed for the forward lunges.

All the movements, except for burpees and forward lunges, have been performed with 4 different external weights. Given that our participants had a similar constitution, the weights that have been chosen were 20, 25, 30 and optionally 35 kg for women and 25, 30, 35 and optionally 37.5 kg for men. We decided to test the same movement with multiple weights in order to be able to study for patterns in the development of power at increasing weight lifted.

These movements have been chosen among the most common ones in functional fitness and taking into consideration that they had to be performed in a laboratory context, without the possibility to use machinery or other big equipment.

During the data collection, the different groups of movements were randomly performed but keeping the same weight progression, from the lightest to the heaviest. The burpees, the only group of movement without external weights, and the forward-lunges, that required a different setup, were always the last movements to be performed for practical reasons in preparing the setup. Moreover, the progression of the burpees was from the least demanding to the most demanding: double step, step up, and jump up.

*2.4 Data post processing*

After the data collection, all the motion capture data were post-processed using the Vicon Nexus 2.15 software: during this step all the reflective markers were used to reconstruct a model of each athlete and his/her movements. Then, these models were further analyzed with the program OpenSim. In here it was possible, through the inverse kinematic and inverse dynamic tools, to get the moment and the angular displacement of every joint. These data were finally used in a Python script to compute angular velocity and joint power. This power was used to check the validity of the power output computed with the IMUs. Due to modeling limitations of the ground contact model on OpenSim, burpees were analyzed only with the force plate data. From the total force recorded (GRF) it was possible to compute the accelerations sustained by the athlete dividing the force by the athlete's mass and removing the gravity. From the acceleration it was then possible to compute the velocity and then, multiplying the force and the velocity, it was possible to obtain the power output. To make this force plate power calculation for burpees as close as possible to the ground truth, the same force plate method was used to compute power output for all the other movements as well and then a linear regression was performed with the motion capture data to find the needed correction between the two methods. Linear regression was chosen as reported also by [30] in their study and as a first way to assess the two curves.

In order to estimate the power output from the three IMUs attached to the athletes and build the power meter that was intended to be created in this study, a Python script was developed. The first step in the script was to use the data recorded during the static calibration to compute the drifts of the sensors. These drifts were then subtracted from the recordings of the movements, performing a bias correction. The following steps were the calculation of the quaternions to estimate the position of each IMU in the 3D space, the extraction of velocity and displacement data from the acceleration data using the cumulative trapezoid method after having removed the gravity acceleration, and, finally, a filtration of all the data. The filtering step was performed using a low-pass Butterworth filter of the 3rd order at 5 Hz for all the data except for the gyroscope data that were filtered using a high-pass Butterworth filter of the 3rd order at 0.05 Hz. At this point, the force data collected with the force plates were imported and filtered with a low-pass Butterworth filter of the 3rd order at 10 Hz. It was then possible to synchronize the IMU and the force plate data, in order to have a shared timestamp for both. This step allowed for a precise identification of the single repetitions for every movement from the video recorded during the test. And, thanks to the synchronization between the videos and the force plate data, it was possible to determine the beginning and the end of every movement in all the recorded data.

After this movement identification it was possible to compute power from the force plates as described earlier. This step was done to have the force plate power output estimation.

The last step was the calculation of the power output for the segments each IMU sensor was accounting for: the wrist was considered to stand for the upper limbs and eventual external weight, the ankle for the lower limbs, and the chest for the trunk. For this calculation, the two approaches described above have been tried, both the one using the $\boldsymbol{F} \cdot \boldsymbol{v}$ formula for instantaneous power (later called force approach), and the one using the difference in mechanical energy (later called energy approach). So, as an example, the power output computed with the wrist worn IMU with the force approach was the following:

$$P_{wrist} = \left(m_{arm} * 2 + m_{ext\_weight}\right) * a_{IMU} * v_{IMU} \tag{4}$$

With the acceleration being the one measured with the IMU applied to the segment and the velocity being the one obtained by integrating the IMU net acceleration of the segment. For the energy approach, the power output computed with the wrist worn IMU was computed with the following formula:

$$P_{wrist} = \frac{\Delta[\frac{1}{2}\left(m_{arm} * 2 + m_{ext_{weight}}\right) * v_{IMU}^2 + \left(m_{arm} * 2 + m_{ext_{weight}}\right) * g * (h_{IMU} - h_{ref})]}{\Delta t} \tag{5}$$

With the velocity being the one obtained by integrating the IMU net acceleration, g being the gravitational acceleration, $h_{IMU}$ being the displacement at any instant of the IMU, derived double integrating IMU net acceleration, and $h_{ref}$ being the IMU displacement in the beginning of the movement.

The reason behind the choice of trying both the formulas (4) and (5) was to see which one could estimate the power output at the best. For both the approaches, the computation was performed considering every single instant recorded, and this method allowed for the calculation of the whole power output curve during each movement. The same computation was performed with the motion capture and the force plate data as well.

These steps have been performed for each of the three sensors applied to the athletes' body and the results have been summed together to get a final power value for each movement.

In order to assess the accuracy of the power meter with a different number of sensors used, the power estimation was also performed just considering the chest and ankle sensors, the chest and wrist sensors or just the chest sensors, since these are the most common combination for athletes that already use a chest strap and/or a smartwatch during their training sessions. The chest and ankle combination was chosen to see if it could be beneficial to have a sensor there while tracking the chest.

After having computed all the power output estimations, it was possible to calculate the mean positive power, the peak positive power and the total positive work, all parameters produced during concentric muscular contraction, and the mean negative power, the peak negative power and the total negative work, the ones produced during eccentric muscular contraction. These parameters were calculated for each single repetition of each single movement, to see how the power production varies between repetitions, but also considering the repetition of the same movement all together, to have a movement level of analysis.

*2.5 Statistical analysis*

The final step in this study was to validate the newly developed power meter pipeline computing the correlations between the two IMU approaches and the motion capture data obtained from OpenSim (the gold standard), in order to discover the best overall approach to the estimation of power in functional fitness. To do this, a Pearson correlation between the whole power output curves computed with the two IMU approaches, and the power curve obtained from the motion capture data was calculated for all the movements except for the burpees, as mentioned before. In this case, the correlation was done between the IMU data and the force plate data. Correlation between the whole curves was chosen to check not only if the average power output or other parameters are well estimated, but also if the whole power estimation is done correctly and thus also the summary statistics elaborated from it. According to what was reported by Jimenez-Olmedo et al. [30]., the results of the correlation were categorized as follows: ≤0.1 (trivial); 0.1-0.3 (low), 0.3-0.5 (moderate), 0.5-0.7 (high), 0.7-0.9 (very high), ≥0.9 (almost perfect).

From a comparison of the power curves, it was possible to calculate a linear regression: this was useful to estimate the relationship between the two curves. The equation derived from this regression is in the form $y = mx + q$, in which the slope $m$ should ideally be 1, to indicate no proportional differences in the two curves, while the intercept $q$ should be ideally close to 0 to indicate no systematic differences. The results of the linear regression were then applied to the IMU output to see if the power estimation improved.

To assess accuracy, precision and agreement of the metrics computed with the IMU-based power meter compared with the gold standard, some statistical evaluations were conducted: Bland-Altman plots were used to measure agreement [30], Root Mean Squared Error (RSME) [28], percentage error and Bland-Altman bias were used to assess the accuracy, while the precision was assessed using the standard deviation of the percentage error and with the Bland-Altman 95% Limit of Agreement (LoA) width. All these tests were performed on both the raw estimation, the one obtained with the power meter, and the corrected estimation, the one in which the linear correlation equation were used to improve the results.

## 3. Results

*3.1 Correlation results*

A total of 147 movements has been recorded and analyzed, considering all the participants and the different movements performed.

At a visual inspection, all the four methods of power estimation (motion capture and force plates as ground truth, IMU with force approach and IMU with energy approach as the new solution) provided very similar results both in the shape of the power output during the different movements but also in the magnitude of the results. An example of the result for thrusters can be seen below (Figure 3):

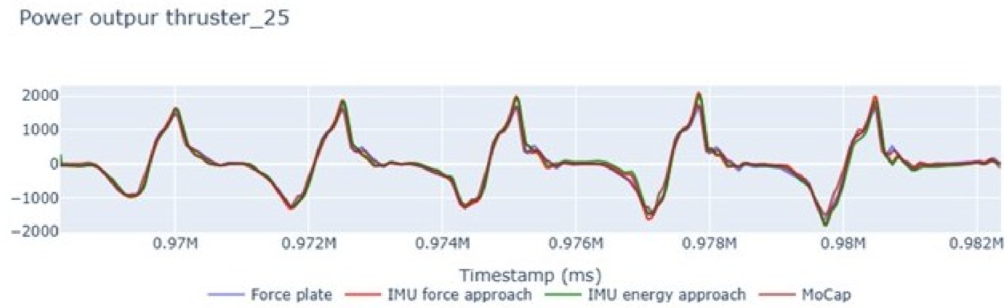



*Figure 3: example of power output curve comparison between the 4 methods*

For each movement, a Pearson correlation and a liner regression were performed. Table 1 and Table 2 show the Pearson correlation coefficients for, respectively, the force approach and the energy approach, both compared with the motion capture system (gold standard) for the 4 different combinations of number of sensors. The two tables only show the overall results obtained considering all the movements together.

*Table 1: overall correlation and linear regression for force approach*

| Number of sensors | Approach | Pearson correlation coefficient |
|---|---|---|
| 3 | Force | 0.908 |
| 2 (chest and ankle) | Force | 0.863 |
| 2 (chest and wrist) | Force | 0.927 |
| 1 | Force | 0.880 |

*Table 2: overall correlation and linear regression for energy approach*

| Number of sensors | Approach | Pearson correlation coefficient |
|---|---|---|
| 3 | Energy | 0.894 |
| 2 (chest and ankle) | Energy | 0.847 |
| 2 (chest and wrist) | Energy | 0.911 |
| 1 | Energy | 0.864 |

The Pearson correlation coefficient shows that the force approach produced estimations that correlate between 0.863 and 0.927 with the ones obtained with the motion capture system. The linear regression showed that there are some proportional differences in the estimation between the two methods. On the other hand, there is a reduced systematic difference.

For the energy approach the Pearson values are similar to the one obtained with the force approach, even though they are always slightly lower being in the range between 0.847 and 0.911. Linear regression showed a smaller proportional difference with the gold standard, while the intercept is always a bit bigger than in the force approach, indicating a bigger systematic difference.

*3.2 Performance evaluation of force approach and energy approach*

To assess the performance of the power meter and of the corrections performed through the linear equations, RMSE, percentage error and Bland-Altman plot were computed both before and after the correction. These statistical tests were used on 6 specific metrics computed from the power curves obtained with the IMU-based power meter and the motion capture system. In order to have a bigger sample, the metrics were computed for every repetition of every movement, having a total of 607 unique repetitions included in the analysis. The 6 metrics, the ones more useful when using a power meter, are the following: average positive and negative power [19], peak positive and negative power [31] and total positive and negative work [19]. In Table 3 and Table 4, respectively for force and energy approach, it is possible to see the overall results of these statistical tests for the forementioned metrics considering all the movements together and the configuration with 3 sensors. The bias and 95% Limit of Agreement (LoA) computed for the Bland-Altman plots were reported in the Tables as well.

*Table 3: validation test for force approach and using 3 sensors*

| **Metric (ref. value)** | **% error raw** | **% error corr** | **RMSE raw (abs, %)** | **RMSE corr (abs, %)** | **Bias raw (95% LoA)** | **Bias corr (95% LoA)** |
|---|---|---|---|---|---|---|
| Avg pos P (483.30 W) | 36.14 ± 35.94 | -9.12 ± 18.75 | 214.88 (44.46%) | 100.68 (20.83%) | 160.90 (-118.25, 440.05) | -44.37 (-221.50, 132.75) |
| Avg neg P (-430.61 W) | 37.16 ± 32.00 | -13.38 ± 15.40 | 194.65 (-45.20%) | 93.51 (-21.72%) | -145.03 (-399.49, 109.43) | 62.16 (-74.76, 199.08) |
| Tot pos W (771.91 J) | 35.89 ± 32.91 | -12.35 ± 14.80 | 375.21 (48.61%) | 177.76 (23.03%) | 281.37 (-205.15, 767.89) | -111.68 (-382.75, 159.39) |
| Tot neg W (-800.39 J) | 32.86 ± 30.05 | -11.79 ± 14.67 | 347.48 (-43.41%) | 178.72 (-22.33%) | -260.47 (-711.25, 190.31) | 112.79 (-158.93, 384.52) |
| Peak pos P (1494.84 W) | 31.61 ± 28.09 | -2.55 ± 20.83 | 595.42 (39.83%) | 377.53 (25.26%) | 459.17 (-283.79, 1202.13) | -79.30 (-802.76, 644.15) |
| Peak neg P (-1356.23 W) | 29.19 ± 39.22 | -4.42 ± 26.46 | 670.46 (-49.44%) | 435.81 (-32.13%) | -383.85 (-1461.28, 693.57) | 97.39 (-735.21, 929.98) |

*Legend: P = power; W = work; Ref. Value = reference (gold standard) mean for each metric; "Raw" refers to uncorrected IMU estimates; "Corr" refers to estimates after applying the correction method.*

*Table 4: validation test for energy approach and using 3 sensors*

| **Metric (ref. value)** | **% error raw** | **% error corr** | **RMSE raw (abs, %)** | **RMSE corr (abs, %)** | **Bias raw (95% LoA)** | **Bias corr (95% LoA)** |
|---|---|---|---|---|---|---|
| Avg pos P (483.30 W) | 31.54 ± 37.11 | -9.72 ± 19.06 | 198.06 (40.98%) | 102.24 (21.16%) | 135.51 (-147.61, 418.63) | -48.99 (-224.89, 126.91) |
| Avg neg P (-430.61 W) | 29.52 ± 29.48 | -13.78 ± 16.99 | 163.88 (-38.06%) | 98.44 (-22.86%) | -112.66 (-345.94, 120.61) | 64.33 (-81.71, 210.37) |
| Tot pos W (771.91 J) | 32.61 ± 32.99 | -11.16 ± 16.50 | 343.03 (44.44%) | 182.79 (23.68%) | 248.37 (-215.36, 712.11) | -109.84 (-396.21, 176.54) |
| Tot neg W (-800.39 J) | 24.16 ± 27.71 | -13.66 ± 15.44 | 286.99 (-35.86%) | 196.36 (-24.53%) | -191.57 (-610.40, 227.27) | 128.83 (-161.63, 419.28) |
| Peak pos P (1494.84 W) | 59.02 ± 101.42 | 22.40 ± 86.31 | 1721.99 (115.20%) | 969.88 (64.88%) | 866.43 (-2050.32, 3783.18) | 254.59 (-1579.56, 2088.93) |
| Peak neg P (-1356.23 W) | 22.64 ± 37.53 | -5.31 ± 25.86 | 594.57 (-43.84%) | 429.38 (-31.66%) | -290.00 (-1307.33, 727.34) | 112.46 (-699.76, 924.67) |

*Legend: P = power; W = work; Ref. Value = reference (gold standard) mean for each metric; "Raw" refers to uncorrected IMU estimates; "Corr" refers to estimates after applying the correction method.*

Looking at Table 3, all the metrics show an overestimation of their output compared with the gold standard. The percentage error is between 29.19% and 37.16% and the standard deviation, thus the variability of the estimation, is quite big. This indicates low precision and low accuracy of the values provided by the raw estimation. The other statistical results support the findings, with an absolute percentage RMSE in the range between 39.83% and 49.44% and with Bland-Altman bias and 95% LoA that are in line with the percentage error and the standard deviation. After the correction, despite the fact that now the power meter underestimates the gold standard, all the metrics show improvements: the percentage error is now in the range between -2.55% and -13.38% and the standard deviation is reduced. This is reflected in the percentage RMSE, where its values are almost halved after the correction for many of the metrics. The bias is reduced as well with a smaller 95% LoAs. The corrected estimation shows, thus, both better precision and accuracy.

For the energy approach (Table 4), the situation is almost the same as the one just seen for the force approach, with the only exception of the peak positive power: with this metric the energy approach is performing quite badly, with a big systematic error and a huge random error, showing very poor precision and accuracy. After the correction the situation improves but it is still not comparable with all the other metrics.

Considering the two Tables together it is possible to see that, except for the peak positive power, energy approach slightly outperforms the force approach before the correction, showing a smaller percentage error (34.2% on average for the force approach and 28.1% for energy approach), a smaller RMSE and a smaller bias. The standard deviation and the 95% LoAs are comparable with the two methods.

After the correction, with nearly all the metrics closer to the gold standard values, the energy approach slightly underperforms the force approach, showing that the correction is more effective with the force approach (mean percentage error being -10.2 for force approach and -10.7% for energy approach, excluding the peak positive power). Nonetheless, the results are quite similar for all the metrics in both the approaches, except for the peak positive power (percentage error -2.55% with force approach and 22.40% with energy approach).

### *3.2 Performance evaluation of different sensor configurations*

In order to check which sensor configuration provides the best results, Table 5 shows a comparison of the performances of the different configurations for the metrics analyzed before. Since, from the previous paragraph, the force approach seems the one that produces slightly better results, the Table only shows the statistics for that approach. Moreover, given that the correction is more effective in improving the power estimation, the Table presents a comparison only for the statistical tests computed on the corrected power output.

*Table 5: comparison of validation test between different sensor configurations*

| Metric (ref. value) | Number of sensors | % error corr | RMSE corr (abs, %) | Bias corr (95% LoA) |
|---|---|---|---|---|
| Avg pos P (483.30 W) | 3 | -9.12 ± 18.75 | 100.68 (20.83%) | -44.37 (-221.50, 132.75) |
| | 2 (chest and ankle) | -17.96 ± 23.17 | 148.66 (30.76%) | -90.89 (-321.42, 139.47) |
| | 2 (chest and wrist) | -9.39 ± 19.82 | 105.74 (21.88%) | -47.89 (-232.67, 136.90) |
| | 1 | -17.16 ± 24.72 | 150.66 (31.17%) | -88.78 (-327.37, 149.81) |
| Avg neg P (-430.61 W) | 3 | -13.38 ± 15.40 | 93.51 (-21.72%) | 62.16 (-74.76, 199.08) |
| | 2 (chest and ankle) | -19.25 ± 17.95 | 124.71 (-28.96%) | 89.80 (-79.81, 259.41) |
| | 2 (chest and wrist) | -12.54 ± 15.80 | 94.53 (-21.95%) | 59.38 (-84.78, 203.53) |
| | 1 | -17.97 ± 18.43 | 121.97 (-28.32%) | 84.63 (-87.51, 256.77) |
| Tot pos W (771.91 J) | 3 | -12.35 ± 14.80 | 177.76 (23.03%) | -111.68 (-382.75, 159.39) |
| | 2 (chest and ankle) | -20.16 ± 19.26 | 259.17 (33.58%) | -179.98 (-545.50, 185.55) |
| | 2 (chest and wrist) | -12.00 ± 16.52 | 182.09 (23.59%) | -107.92 (-395.38, 179.55) |
| | 1 | -18.87 ± 21.42 | 259.71 (33.64%) | -169.65 (-555.07, 215.78) |
| Tot neg W (-800.39 J) | 3 | -11.79 ± 14.67 | 178.72 (-22.33%) | 112.79 (-158.93, 384.52) |
| | 2 (chest and ankle) | -18.93 ± 19.04 | 256.83 (-32.09%) | 179.96 (-179.17, 539.10) |
| | 2 (chest and wrist) | -11.36 ± 15.70 | 179.27 (-22.40%) | 108.45 (-171.33, 388.23) |
| | 1 | -17.76 ± 20.64 | 256.23 (-32.01%) | 170.12 (-205.44, 545.67) |
| Peak pos P (1494.84 W) | 3 | -2.55 ± 20.83 | 377.53 (25.26%) | -79.30 (-802.76, 644.15) |
| | 2 (chest and ankle) | -3.62 ± 25.35 | 444.44 (29.73%) | -98.14 (-947.75, 751.46) |
| | 2 (chest and wrist) | 1.66 ± 19.24 | 320.08 (21.41%) | -10.35 (-637.38, 616.68) |
| | 1 | -0.59 ± 23.19 | 377.99 (25.29%) | -46.07 (-781.41, 689.27) |
| Peak neg P (-1356.23 W) | 3 | -4.42 ± 26.46 | 435.81 (-32.13%) | 97.39 (-735.21, 929.98) |
| | 2 (chest and ankle) | 2.64 ± 34.93 | 487.57 (-35.95%) | 21.82 (-932.87, 976.51) |
| | 2 (chest and wrist) | -0.38 ± 28.21 | 427.48 (-31.52%) | 48.33 (-784.16, 880.82) |
| | 1 | 5.72 ± 34.99 | 474.22 (-24.97%) | -19.56 (-948.23, 909.11) |

*Legend: P = power; W = work; Ref. Value = reference (gold standard) mean for each metric; "Raw" refers to uncorrected IMU estimates; "Corr" refers to estimates after applying the correction method.*

For all the metrics but the peak powers, it is possible to see that the performances can be divided into two groups: the first group is the configuration with 3 sensors and the one with the 2 sensors at the chest and the wrist, while the second groups has the configuration with 2 sensors at the chest and at the ankle and the single sensor at the chest. Both the configuration in the first group have an absolute percentage RMSE between 20.83% and 23.59%, excluding the peak powers and both the configurations in the second group have an absolute percentage RMSE between 28.32% and 33.64%, excluding the peak powers. Thus, among these two groups there is a difference around 8%-10% in the percentage RMSE, with the first one providing better estimations. Comparing the configuration in the first group, the 3-sensor configuration and the one with 2 sensors applied to chest and wrist, the latter provides the best results, with a mean percentage error of -7.3% against -8.9% of the first.

The only metrics that don't fit in the two groups division are the peak powers: their performances are more mixed between the different configurations, with the single sensor providing the best estimation for the peak negative power (systematic error -19.56 W and percentage RMSE -24.97%) while the 2 sensors worn at the chest and at the wrist provide the best estimation for the peak positive power (systematic error -10.35 W and percentage RMSE 21.41%).

## 4. Discussion

This study aimed to analyze the concurrent validity of a newly developed IMU-based power meter when compared to a combination of a motion capture system and force plates for the estimation of power output in functional fitness athletes. The power meter has been designed to be based on 3 IMUs, one worn at chest, one at wrist and one at ankle. Two algorithms with different approaches have been developed to estimate the power output using the IMU data as input. The first algorithm is based on force approach, in which power (P) is estimated by multiplying force (F) and velocity (v) according to the formula $P = \boldsymbol{F} \cdot \boldsymbol{v}$ (an example usage can be seen in formula (4) in the material and methods section); the second algorithm is based on energy approach, in which power is derived from a change in the mechanical energy of the athlete over time (an example usage can be seen in formula (5) in the material and methods section).

At a visual inspection, all the four methods of power estimation (motion capture and force plates as ground truth, IMU with force approach and IMU with energy approach as the new solution) provided very similar results both in the shape of the power output during the different movements but also in the magnitude of the results.

To confirm and quantify these similarities and to be able to validate the IMU-based power meter, some statistical analyses were conducted. Thanks to these analyses, it was possible to determine the best approach (force or energy) to estimate power output and the best combination of sensors. Indeed, the configuration with 3 IMUs can be suitable for a laboratory but, in a gym, it is difficult that athletes have so many sensors. They usually have a sensor applied on the chest and, sometimes, a smartwatch that can help. Thus, an investigation about different number of sensors used and worn in different locations has been made to assess the reliability of their results. The possible alternatives to the 3 sensors configuration were two sensors (worn at the chest and the wrist or, in another configuration, at the chest and the ankle) or just one sensor (placed on the chest).

The first analysis conducted to validate the power meter was the calculation of Pearson correlation coefficient, to check if the estimated power and the one computed with the gold standard were well correlated. All the combinations of number of sensors and approaches used for the power estimation showed a very high to almost perfect correlation (0.847-0.927). The results obtained are in line or slightly better than the ones obtained by Jimenez-Olmedo et al. [30], who reported a correlation between IMU and motion capture of 0.847 for IMU worn at the pelvis level. Other studies, investigating the correlation of IMU attached to the torso, found values that were ranging from 0.72 to 0.95. [32], [33], [34]. The best correlation (0.927) was produced with 2 sensors placed at the chest and at the wrist and using the force approach. On the other hand, the lowest correlation (0.847) was obtained with 2 sensors worn at the chest and at the ankle and using the energy approach.

There can be two different factors that can affect the estimation with the configuration with 2 sensors placed at the chest and the ankle with energy approach. The first one is that the energy approach estimation is a little bit more flattened compared to the one produced by the force approach. This fact can be seen both visually inspecting the curves, where the energy estimation is always smaller than the force one, but also looking at the metrics estimated from the power output curve. Except for the peak positive power, all the other metrics present smaller values compared to the ones estimated with the force approach, considering the raw estimation obtained by the power meter, i.e. without correcting the estimation with the linear equation. Thus, a slightly more flattened curve can correlate slightly worse with the gold standard. A second factor can be connected with the sensor configuration: indeed, many movements were involving the upper limbs and the absence of a sensor dedicated to the arms can be a problem in the estimation of the power produced in that part of the body. Even more so if one thinks that the arms were moving external weights, thus producing a considerable amount of power. This, though, does not explain why the worse configuration is the one with 2 sensors (chest and ankle) and not the one with just one sensor: the most likely reason behind this, and also behind the fact that 2 sensors (chest and wrist) are better than 3, is that the sensor placed at the ankle produces a misleading power output instead of valuable one, thus making slightly worse the final power output estimation when the ankle is monitored. This aligns with a finding reported by Tan et al. [35] that a single misplaced IMU can reduce the ground reaction force estimation accuracy by up to 1.1%. Considering that, with IMUs, the force in is estimated from the acceleration, it is easy to understand how this accuracy reduction can affect the power estimation as well.

A second analysis of the power meter was meant for measuring its accuracy, precision and agreement with the data obtained with the motion capture system. Confronting the results of percentage error, RMSE, bias and 95% LoAs both for the force approach and the energy approach, it is possible to note an interesting thing: when computing those statistical tests on the raw power output estimation, i.e. on the estimation given by the power meter without the correction of the linear regression equation, the energy approach provides better outcomes in 5 of the 6 metrics analyzed, with the exception of the peak positive power. Indeed, the absolute percentage error with energy approach for these five metrics is in the range 22.64%-32.61%, while for the same metrics the force approach estimation is in the range 29.19%-37.16%. For the peak positive power, on the other hand, the force approach gives much better estimations, with the percentage error for the force approach being 31.61% and 59.02% for the energy approach. The other statistical tests show similar results. No other studies were found to test power estimation with the same movements as the present one, so the only comparison can be made with studies estimating the power output during a countermovement jump test. Rantalainen et al. [33] reported to have a percentage error of 25.8% between mean concentric power estimated with an IMU compared with power estimated with a force plate during countermovement jumps. The present study is performing a bit worse, 36.14% with the force approach and 31.54% with the energy approach. However, as already mentioned, the movements analyzed are different, so a direct comparison is difficult.

Applying the linear correction the power estimation improves for all the metrics using the force approach (absolute percentage error in the range 2.55%-13.38%) and for almost all the metrics using the energy approach (absolute percentage error in the range 5.31%-22.40%). The variability, expressed by the standard deviation and by the 95% LoAs, is also reduced nearly for all the metrics. After the correction the force approach appears to be slightly better than the energy approach.

The only metric that improves but is still not in line with all the other is the positive peak power estimated using the energy approach. To understand why this happens it is useful to check the Bland-Altman plots. In Figure 4 it is possible to see the Bland-Altman plots for the peak positive power in the configuration with the 3 sensors and estimated using energy approach.

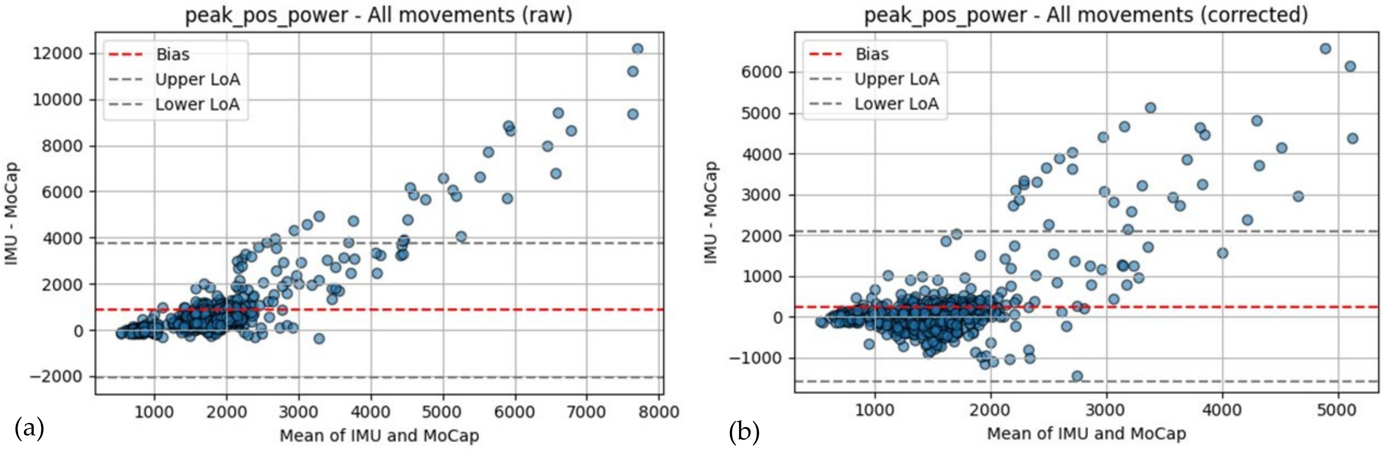



*Figure 4: Bland-Altman plot for peak positive power estimated with energy approach before (a) and after the correction (b)*

In Figure 5 the same metric is represented for the same configuration of sensors but showing the result of the force approach.

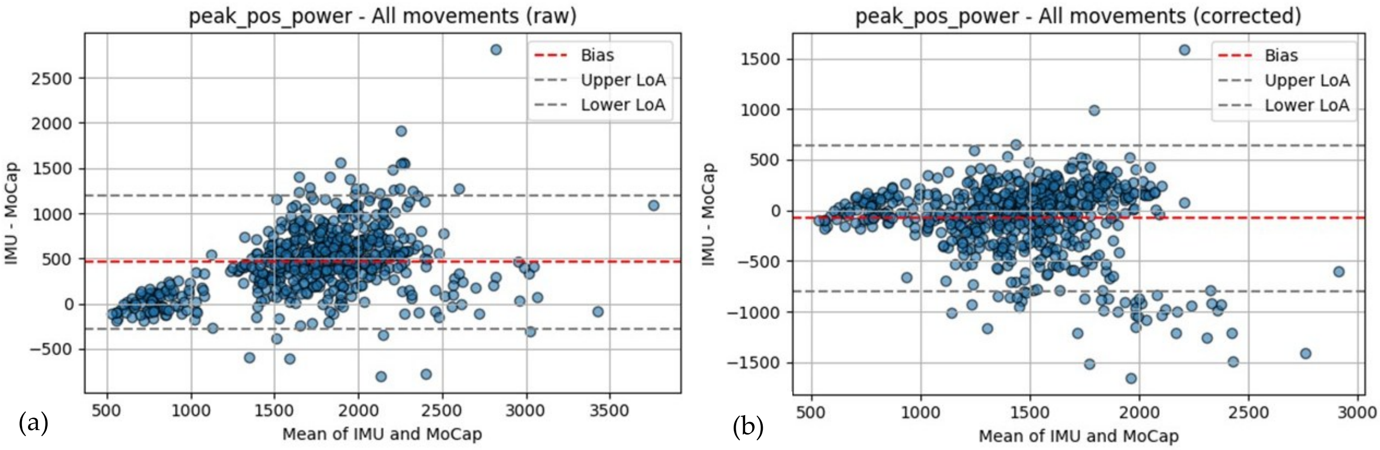



*Figure 5: Bland-Altman plot for peak positive power estimated with force approach before (a) and after the correction (b)*

It is evident that, in the energy approach there are some values that have a trend similar to the equation y = x and that are influencing the results a lot.

To understand where the problem is, Table 7 shows the statistical results of peak positive power divided per single movement.

*Table 7: peak positive power with energy approach analysis*

| Movement (ref. value) | % error raw | % error corr | RMSE raw (abs, %) | RMSE corr (abs, %) | Bias raw (95% LoA) | Bias corr (95% LoA) |
|---|---|---|---|---|---|---|
| Burpee (1996.11 W) | 124.99 ± 149.33 | 38.39 ± 88.24 | 3713.15 (186.02%) | 1805.38 (90.44%) | 2413.13 (-3069.83, 7932.09) | 730.06 (-2506.27, 3966.39) |
| Clean and jerk (1500.06 W) | 29.62 ± 29.92 | -5.91 ± 21.44 | 586.78 (39.12%) | 319.24 (21.28%) | 426.11 (-364.57, 1216.80) | -97.94 (-693.49, 497.60) |
| Lunge (793.35 W) | 89.66 ± 183.05 | 94.18 ± 186.74 | 1485.79 (187.28%) | 1524.55 (192.17%) | 649.01 (-1970.61, 3268.63) | 683.24 (-1988.00, 3354.48) |
| Snatch (1412.26 W) | 53.33 ± 24.90 | 5.12 ± 16.93 | 777.56 (55.06%) | 236.87 (16.77%) | 720.24 (145.92, 1294.55) | 52.66 (-399.99, 505.31) |
| Thruster (1616.09 W) | 25.55 ± 15.40 | 12.10 ± 13.17 | 441.53 (27.32%) | 253.26 (15.67%) | 398.34 (25.05, 771.64) | 185.24 (-153.25, 523.74) |

*Legend: Ref. Value = reference (gold standard) mean for each movement; "Raw" refers to uncorrected IMU estimates; "Corr" refers to estimates after applying the correction method.*

From the Table above it is evident that the worst results are obtained with burpees and lunges, while the other movements show performances that are in line with the general results of the other metrics seen above. Indeed, the percentage RMSE is in the range 15.65%-21.28% for the clean and jerk, snatches and thrusters, and this is comparable with the percentage RMSE of the same metric but estimated with the force approach on all the movements (25.26%). Plotting Blend-Altman plots for all the movements confirms this finding (Figure 6 for the plots of burpees and lunges, Figure 7 for the plots of the other movements).

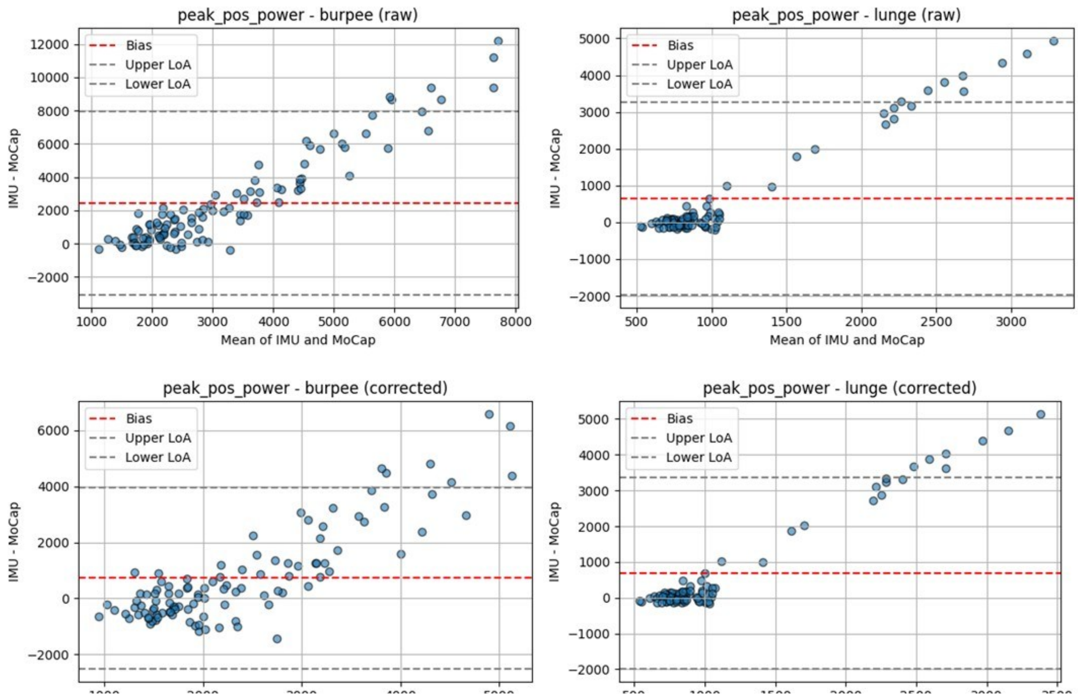



*Figure 6: Bland-Altman plots for peak positive power estimated with energy approach for burpees and lunges. The first row shows the results before correction, the second row after the correction*

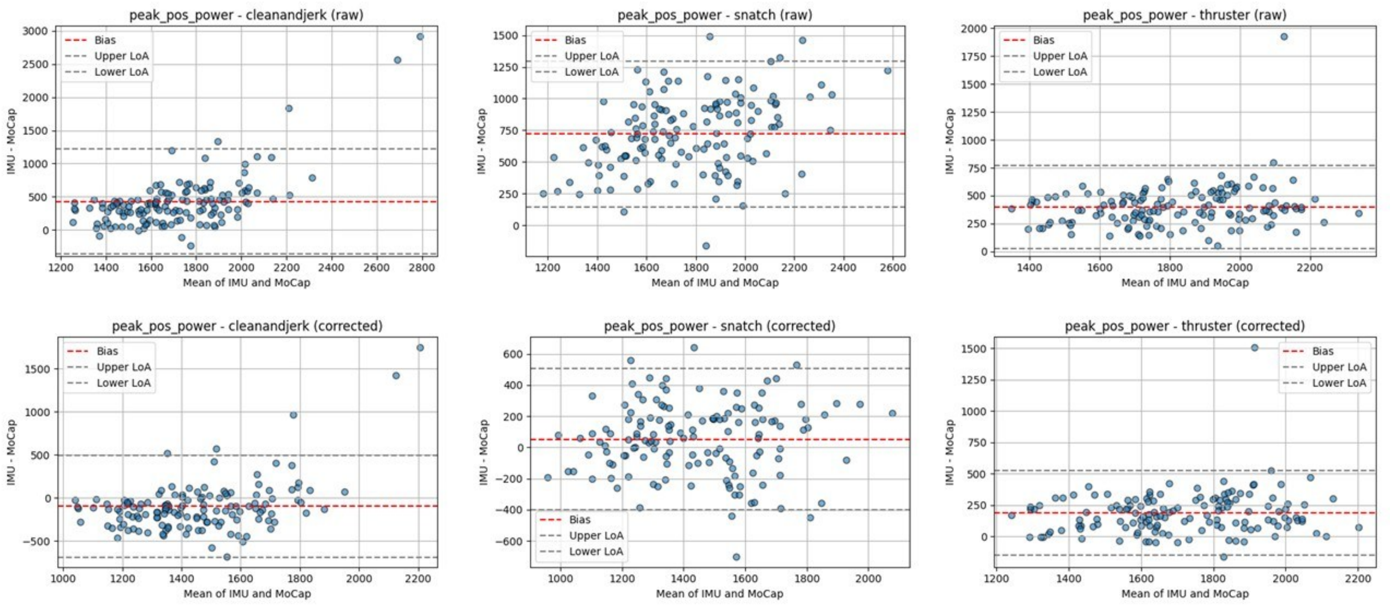



*Figure 7: Bland-Altman plots for peak positive power estimated with energy approach for clean and jerks, snatches and thrusters. The first row shows the results before correction, the second row after the correction*

The problem in estimating peak positive power with burpees and lunges is due to the fact that, in order to minimize the drifting of the acceleration during the recordings, thus reducing the possible estimation error due to double integration of the signal over time to compute displacement, several corrections were used during the data processing. These were the static calibration, the removal of the offset during the integration process and the reinitialization of the reference height and mechanical energy for computing the displacement at the beginning of every movement and every repetition. This last point is the one that is more likely to cause the problem. Indeed, this reinitialization paired with the fact that burpees and lunges were the only movements where the transition between a repetition and the other didn't present an instant where the mechanical energy could be assumed near zero, made that, at the beginning of some repetitions in those two movements, some jumps in the mechanical energy could have happened. And these jumps in mechanical energy are the ones that, then, produce spikes in the power output that can cause this poor performance in the estimation of the peak positive power.

To try to solve this problem a solution can be to avoid the reinitialization of reference height and mechanical energy in the middle of those two movements, trying to reduce the drift that might eventually arise in other ways. Another possibility is to implement a strategy in the searching of peaks that detect possible outliers and exclude them from the possible peak candidates.

Finally, speaking of the best configuration of sensors to estimate the power output, from the data collected it appears that the 3 sensors and the 2 sensors placed at the chest and at the wrist are quite close in many metrics, with the configuration with two sensors that globally outperform the other. Indeed, it has a mean percentage error of -7.3% against the -8.9% of the version with 3 sensors. This value, despite being significantly different from the gold standard ($p < 0.001$), can still be acceptable as first estimation if the people that are going to use this device are aware of it. The pros deriving from being able to assess power output from athletes, even with a bit of underestimation, are bigger than the risks deriving from a lack of assessment or an assessment with a tool that provides wrong results, especially if the underestimation is taken into account. Indeed, the risks deriving from a lack of performance assessment or from wrong power output estimations are that coaches and athletes take decisions for some training sessions that could be harmful to athletes resulting in an injury or, less severe, in a reduction of performance. This is also highlighted by Gabbett [36], that shows that non-contact injuries are due to inappropriate training programs. This can be seen both as an ethical and social problem, as the athletes need to suspend training, spend money on healthcare to recover from the injury, without considering their physical and psychological suffering.

So, with this study it was possible to bridge the existing gap between endurance sports and functional fitness in the way the performances of the athletes are measured. It is now possible to compare different parts of the same training program to have a general estimation of the load sustained by the athlete. Moreover, this allows comparison with other sports as well, showing how much power is developed while doing a clean and jerk or a burpee as well as how much of it is developed while doing an all-out sprint in a road cycling race or the effort required to climb a famous mountain in one of the cycling grand tours' stages. For example, Sanders et al. [37] reported that the average power output of 20 male professional cyclists analyzed in their races for 4 years (mean duration 285 min) was 216 W (3.0 W/kg), while the same outcome measured for 10 professional female cyclists on a mean duration of 194 min was 167 W (2.8 W/kg). Considering a more intense and shorter effort, Vogt et al. [38] reported that the average power output of six male professional cyclists during an uphill time trial stage was 392 W (5.5 W/kg). These data can be confronted with the results of this study, in which all the athletes expressed an average power of 483.30 W (6.57 W/kg) considering all the movements they performed. So, the power values of functional fitness athletes can be even higher than the ones expressed by professional cyclists, even though it must be taken into consideration the much shorter duration of the analyzed movements in this study and the fact that functional fitness athletes develop power with the whole body while the cyclists do that mostly with their legs.

### *4.1 Limitations*

Despite all the efforts in trying to carry out a study with as little biases as possible, some limitations are still affecting the present study. The first limitation is the reduced dimension of the sample analyzed, so it is difficult to draw statistically valid conclusions and also to generalize the findings to a broader group of athletes. Moreover, the fact that the data collection was held in a laboratory limited the number of available exercises and of the available external weight that could be used. Another thing to consider is that with all the markers and the IMUs placed on athletes body and the fact that they had to perform all the movements on the force plates, they couldn't perform their usual techniques for the movements we had in our test, hence the power output could be not reflective of the real power expressed in a gym context. The researchers tried to limit the influence of this limitation as much as possible, like placing soft markers where athletes were making contact with the barbell to allow them to perform as if there were no markers.

### *4.2 Future directions*

As already mentioned earlier, trying to minimize the drift of the energy method without creating spikes in the power estimation could improve the energy approach a lot. Several strategies can be tried here: one strategy could be to compute repetition power starting from the whole movement power estimation. In this way, the signal is continuous, thus avoiding the spikes at the beginning of the repetitions. Another possibility could be the implementation of an outlier detection step that removes the artificial spikes before the identification of the peak power.

In the future it could also be interesting to expand the sample to see if the results discussed in this paper are confirmed. Moreover, it would be interesting to find a good ground truth, instead of the motion capture system, to be used directly in the gyms, where the athletes can be tested on many other movements and with the possibility of performing the gestures at their best, without interference due to laboratory constraints. This could help a lot in generalizing the results and capturing data that are closer to the ones produced during the real potential use of such a device.

## 5. Conclusion

The aim of this work was to develop and validate a power meter that can be used by functional fitness athletes. To build a device that can be reliable without interfering with athletes' gestures, the idea was to use some inertial measurement units (IMUs) applied to the chest, the wrist and the ankle of the athletes. This configuration was chosen after reviewing literature for the best IMU position on the body and with the idea that a sensor applied to the trunk, one on the upper limb and one on the lower limb were the best and most comprehensive configuration for power estimation.

To validate the power meter some functional fitness movements were performed in a biomechanic laboratory while the athletes were wearing the three IMUs and reflective markers to be used in combination with a motion capture system (gold standard). The data collected by the IMUs were then used to estimate power output and this was compared with the power output calculated with the data collected by the gold standard. The IMU-based power meter was built with two different approaches that followed by the literature analysis conducted before the beginning of the study: the first approach used was the force approach, in which power (P) is estimated multiplying force (F) and velocity (v) according to the formula $P = \boldsymbol{F} \cdot \boldsymbol{v}$; the second approach was the energy approach, in which the power was derived from a change in the mechanical energy of the athlete.

The results of the validation study suggested that the force approach provided the best correlation with the motion capture and, after the correction of the power output estimation with the linear equation found with a linear regression, the force approach was the one providing closer estimations to the gold standard, despite the difference with the gold standard was still statistically significant. When investigating the performance of different configurations of sensors, the configuration with just one sensor applied to the chest and one applied to the wrist showed the best overall results, even if it was quite close to the configuration with 3 sensors.

## 6. Patents

This work resulted in a patent claim in the US for the company WodMotions.

**Funding:** The authors are employees of WodMotions and they are paid by the company.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Royal Institute of Technology (KTH).

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

**Data Availability Statement:** The data presented in this study are not available as they are confidential.

**Acknowledgments:** The researchers would like to thank all the athletes who took part in the data collection and the staff at the Promobilia MoveAbility Lab in KTH for their equipment and for their assistance.

**Conflicts of Interest:** The authors declare themselves to be part of the WodMotions team.

## Abbreviations

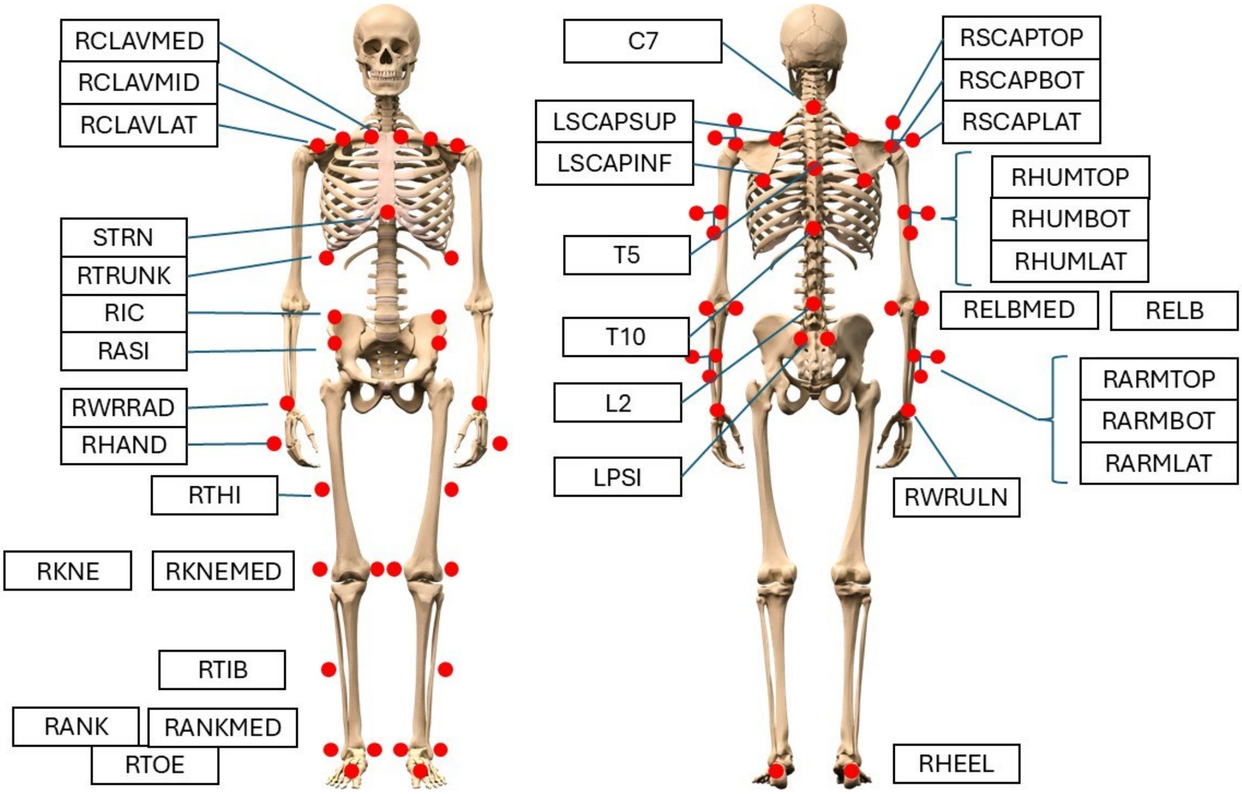
The following abbreviations are used in this manuscript:

| | |
|---|---|
| 1RM | One Repetition Maximum |
| FTP | Functional Threshold Power |
| GRF | Ground Reaction Force |
| IMU | Inertial Measurement Unit |
| LoA | Limit of Agreement |
| RMSE | Root Mean Square Error |

## Appendix A

*Appendix A1*

Picture showing the disposition of the 67 reflective markers used in this study, according to a modified CGM 2 model



## References


[1] L. Passfield, J. Hopker, S. Jobson, D. Friel and M. Zabala, "Knowledge is power: Issue of measuring training and performance in cycling," *Journal of Sports Science,* 2016.

[2] V. de Vette, D. Veeger and M. van Dijk, "Using wearable sensors to estimate mechanical power output in cyclical sports other than cycling - a review," *Sensors,* pp. 23, 50, 2023.

[3] H. Allen, A. Coggan and S. McGregor, Training and racing with a power meter, 3rd Ed., VeloPress, 2019.

[4] D. Jaén-Carrillo, L. Roche-Seruendo, A. Carton-Llorente, R. Ramirez-Campillo and F. Garcia-Pinillos, "Mechanical power in endurance running: A scoping review on sensors for power output estimation during running," *Sensors,* 2020.

[5] G. Mangine and T. Seay, "Quantifying CrossFit: Potential solutions for monitoring multimodal workloads and identifying training targets," *Frontiers,* 2022.

[6] Y. Feito, K. Heinrich, S. Butcher and W. Poston, "High-intensity functional training (HIFT): Definition and research implications for improved fitness," *Sports,* vol. 6, no. 76, 2018.

[7] CrossFit, "CrossFit," [Online]. Available: https://www.crossfit.com/essentials/magic-the-methodology. [Accessed 19 June 2025].

[8] G. Glassman, "Understanding Crossfit," *The Crossfit journal,* vol. 56, no. 1, 2007.

[9] J. Dexheimer, E. Schroeder, B. Sawyer, R. Pettitt, A. Aguinaldo and W. Torrence, "Physiological performance measures as indicators of CrossFit performance," *Sports,* vol. 7, no. 93, 2019.

[10] R. Martinez-Gomez, P. Valenzuela, L. Alejo, J. Gil-Cabrera, A. Montalvo-Perez and E. e. a. Talavera, "Physiological predictors of competition performance in CrossFit athletes," *International journal of environment research and public health,* vol. 17, no. 3699, 2020.

[11] G. Mangine, Y. Feito, J. Tankersley, J. McDougle and B. Kliszczewicz, "Workout pacing predictors of CrossFit open performance: a pilot study," *Journal of human kinetics,* vol. 78, pp. 89-100, 2021.

[12] B. Kliszczewicz, G. Mangine, E. Zeitz, J. Dexheimer and J. Tankersley, "Pacing strategies in a nine-minute high intensity functional training competition workout including deadlifts, handstand pushups, and handstand walks," in *National conference*, Orlando, FL, 2021.

[13] G. Mangine, J. Dexheimer, E. Zeitz, J. Tankersley and B. Kliszczewicz, "Pacing strategies for women in a 20 minute high intensity functional training competition workout containing muscle ups, rowing, and wall balls," in *National conference*, Orlando, FL, 2021.

[14] E. Zeitz, G. Mangine, J. Dexheimer, J. Tankersley and B. Kliszczewicz, "Pacing strategies in a 20-minute high-intensity functional training competition workout including box jumps, clean and jerks, and single leg squats," in *National conference*, Orlando, FL, 2021.

[15] J. Dexheimer, G. Mangine, B. Z. E. Kliszczewicz and J. Tankersley, "Differences between males and females on the same high-intensity functional training competition workout with different prescribed loads," in *National conference*, Orlando, FL, 2021.

[16] G. Glassman, "CrossFit Training Guide Level 1," *The CrossFit Journal,* 2011.

[17] E. van der Kruk, F. van der Helm, H. Veeger and A. Schwab, "Power in sports: A literature review on the application, assumption, and terminology of mechanical power in sport research," *Journal of Biomechanics,* 2018.

[18] G. Noffal and S. Lynn, "Biomechanics of power in sport," *Strength and conditioning journal,* vol. 34, no. 6, pp. 22-24, 2012.

[19] J. Garhammer, "A review of power output studies of Olympic and powerlifting: Methodology, performance prediction, and evaluation tests," *Journal of Strength and Conditioning,* pp. 76-89, 1993.

[20] H. Rangala, S. Samaraweera, K. Sandaruwan, T. Weerasinghe and D. Ranasinghe, "Athlete body power and strength estimation using skeleton point cloud," *Journal of the national science foundation of Sri Lanka,* pp. 481-492, 2024.

[21] V. Camomilla, E. Bergamini, S. Fantozzi and G. vannozzi, "Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: a systematic review," *Sensors,* vol. 18, no. 873, 2018.

[22] J. Taborri, J. Keogh, A. Kos, A. Santuz, A. Umek and C. e. a. Urbanczyk, "Sport biomechanics applications using inertial, force, and EMG sensors: a literature overview," *Applied bionics and biomechanics,* 2020.

[23] I. Poitras, D. F., M. Bielmann, A. Campeau-Lecours, C. Mercier, L. Bouyer and J. Roy, "Validity and reliability of wearable sensors for joint angle estimation: a systematic review," *Sensors,* 2019.

[24] V. Rahn, L. Zhou, E. Klieme and A. B., "Optimal sensor placement for human activity recognition with a minimal smartphone-IMU setup," in *10th International Conference on Sensor Networks*, 2021.

[25] C. Walmsley, S. Williams, T. Grisbrook, C. Elliot, C. Imms and A. Campbell, "Measurement of upper limb range of motion using wearable sensors: a systematic review," *Sports Medicine,* 2018.

[26] W. Niswander, W. Wang and K. Kontson, "Optimization of IMU sensor placement for the measurement of lower limb joint kinematics," *Sensors,* 2020.

[27] A. Cuesta-Vargas, A. Galan-Mercant and J. Williams, "The use of inertial sensors system for human motion analysis," *Physical Therapy Reviews,* vol. 15, no. 6, pp. 462-473, 2010.

[28] H. Snarberg, "Development and validation of a wearable solution for vertical jump tests with immediate feedback," 2024.

[29] M. Mohammadian, H. Mozayyany, S. B. Koudakani and S. Maloney, "Validity and reliability of torso- versus waist-worn inertial measurement units in the assessment of vertical jumps," *Journal of biomechanics,* vol. 176, 2024.

[30] J. Jimenez-Olmedo, B. Pueo, J. Mossi and L. Villalon-Gash, "Concurrent Validity of the Inertial Measurement Unit Vmaxpro in Vertical Jump Estimation," *Applied Sciences,* 2023.

[31] M. Soriano, K. Kipp, J. Lake, T. Suchomel, P. S. D. B. M. Marin and P. Comfort, "Mechanical power production assessment during weightlifting exercises. A systematic review," *Sports biomechanics,* 2023.

[32] R. Spangler, T. Rantalainen, P. Gastin and D. Wundersitz, "Inertial Sensors are a Valid Tool to Detect and Consistently Quantify Jumping," *International Journal of Sports Medicine,* 2018.

[33] T. Rantalainen, P. Gastin, R. Spangler and D. Wundersitz, "Concurrent validity and reliability of torso-worn inertial measurement unit for jump power and height estimation," *Journal of Sports Sciences,* 2018.

[34] J. Pino-Ortega, J. Garcia-Rubio and S. Ibanez, "Validity and reliability of the WIMU inertial device for assessment of the vertical jump," *PeerJ,* 2018.

[35] T. Tan, D. Chiasson, H. Hai and S. P.B., "Influence of IMU position and orientation placement errors on ground reaction force estimation," *Elsevier,* 2019.

[36] T. Gabbett, "The training-injury prevntion paradox: should athletes be training smarter and harder?," *British Journal of Sport medicine,* 2016.

[37] D. Sanders, T. van Erp and J. de Koning, "Intensity and Local Characteristics of Professional Road Cycling: Differences Between Men's and Women's Races," *International Journal of Sports Physiology and Performance,* 2019.

[38] S. Vogt, L. Heinrich, Y. Schumacher, A. Blum, K. Roecker, H. Dickhuth and A. Schmid, "Power Output duringStage Racing in Professional Road Cycling," *Medicine & Science in Sports & Exercise,* 2006.

TRITA – CBH-GRU-2025:041
Stockholm, Sweden 2025

www.kth.se